INEQUALITY RESTRICTED AND PRE-TEST ESTIMATION IN A MIS-SPECIFIED ECONOMETRIC MODEL

A THESIS PRESENTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT

THE UNIVERSITY OF CANTERBURY,

CHRISTCHURCH,

NEW ZEALAND

BY

ALAN TZE-KIN WAN

MARCH 1993

TO MY PARENTS

CONTENTS

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

LIST OF TABLES AND FIGURES

LIST OF TABLES AND FIGURES (CONTINUED)

LIST OF TABLES AND FIGURES (CONTINUED)

LIST OF TABLES AND FIGURES (CONTINUED)

ABSTRACT

This thesis is concerned with the finite sample properties of some estimators of the unknown parameters in a linear model which is (possibly) mis-specified through the exclusion of relevant regressors. We assume that in addition to sample information, prior information regarding the unknown parameters is available in the form of a linear inequality constraint imposed on the regression coefficients. The combination of this type of prior information and sample information in specifying the corresponding statistical model leads to what has been identified in the literature as the inequality restricted estimator. If the statistical significance of the inequality constraint is tested prior to the estimation process, then the estimator thereby generated is called the inequality pre-test estimator.

The properties of these estimators of the coefficient vector in a properly specified model have been examined rather thoroughly in the literature. In this thesis, we extend the results reported in the literature to the case where the underlying regression model is underfitted. We also investigate the sampling performance of the corresponding estimators for the model's disturbance variance, as well as the choice of an optimal size for the pre-test. The general background and motivation for this study are given in Chapter 1.

Much of the earlier research on inequality restricted and pre-test estimation are built on results from studies that assume that the prior information is in the form of linear equality restrictions. We survey the relevant literature in this area in Chapter 2. Chapter 3 reviews the literature on inequality restricted and pre-test estimation. We focus on this problem in the context of the standard linear model with a single linear

inequality constraint on the coefficient vector, as this is directly related to the theme of this thesis.

In Chapter 4, we derive and evaluate the risk, under quadratic loss, of the inequality restricted and pre-test estimators for the regression prediction vector in an underfitted model. This analysis takes the established literature further by allowing for mis-specification in the regressor matrix. We consider the risk of the prediction vector, rather than the coefficient vector itself, so that our results are data independent. The risk functions of the corresponding estimators for the regression disturbance variance in the properly specified and underfitted models are derived in Chapters 6 and 7 respectively.

As in the case where the prior information exists as linear equality restrictions, our results show that when the model is underfitted, the use of valid prior information does not necessarily guarantee a reduction in risk. This result holds for the estimation of both the prediction vector and the scale parameter. When one is estimating the regression disturbance variance, with an appropriate choice of test size, the inequality pre-test estimator can uniformly dominate the estimator that uses sample information only. We also find that the risk functions of the estimators of the error variance are affected more by mis-specification than are the corresponding predictive risks.

In the case where no strictly dominating estimator exists, the question of the choice of an optimal critical value of the pre-test remains. Chapters 5 and 8 explore this issue when one is estimating the prediction vector and scale parameter respectively. We find that most of our results concur qualitatively with those reported in the literature when the prior information exists as exact equality restrictions.

Chapter 9 contains some concluding remarks and a summary of the major results obtained in earlier chapters. We also outline some possible future research topics in this general area.

ACKNOWLEDGEMENTS

First and foremost, I owe an enormous debt to my supervisors, Professor David Giles and Dr. Judith Giles, who have most patiently guided me through innumerable discussions, helped me to clarify my thoughts at each stage, offered constructive criticisms on various drafts of this thesis, as well as providing constant encouragement. It is truly the case that my work would never have progressed so smoothly without their indefatigable guidance. I also thank Professor Giles who initially encouraged my intent to pursue a doctoral degree, and subsequently arranged my access to the department's Ph.D programme.

In addition, I am grateful to the support of Mr. Alan Woodfield and Dr. Judith Giles (former and present Head of the Department of Economics) for providing me with part time teaching assignments, superb office accomodation and computing facilities. Rev. Don Wilson kindly appointed me as a Residential Assistant at the University's Halls of Residence, which eased the financial burden of writing this thesis.

Parts of this thesis were presented at the Australasian Meeting of the Econometric Society, Melbourne, July 1992, and also at the 5th Ph.D Conference of Economics and Business, Canberra, November 1992. I am extremely grateful to the Department of Economics at Canterbury University for providing me with the financial support to attend these conferences. Thanks are also due to Professor Bill Griffiths (University of New England, Australia) for his helpful comments at the latter conference.

My thanks are also offered to the friendly co-operation of various people, who generously supplied some of their unpublished papers and brought my attention to the publications that I was not previously aware of. In this regard, I would like to thank Professors Bill Farebrother (University of Manchester, U.K.), Hikaru Hasegawa (Hokkaido University, Japan) and Kazuhiro

Ohtani (Kobe University, Japan).

Several others have offered insights at various junctures. They include Kevin Albertson (University of Salford, U.K.), Leigh Roberts (Victoria University of Wellington), Kun-Lam Shea (University of Hong Kong), John Small and Keith Wignall. I also benefitted from conversations with Professor Kazuhiro Ohtani during his visit to Canterbury in July 1992. To all of them, I offer my thanks and gratitude.

Numerous other friends and peers have provided encouragement during the process of writing this thesis, for which I thank them all. In particular, this thesis would never have been completed without the support of my girlfriend, Felicity Lye Fun, who has borne with me so sympathetically despite our distance apart. Her understanding has frequently helped to uplift my morale in times of emotional distress. Finally, my most heartfelt thanks are to my parents, for their ample love and constant prayers. I wish to dedicate this thesis to my parents, in appreciation of their love and support throughout all these years.

CHAPTER 1
INTRODUCTION

1.1 GENERAL BACKGROUND AND MOTIVATION

The classical approach to econometric modelling is firmly based on the frequentist principle of statistical inference. The data generating process that gives rise to the observed data is represented by a population regression model which is postulated *a priori*. The observed data are merely a realisation of the data generating process described by the population regression model. Probability is interpreted by frequentists as the long run relative frequency of an event. This interpretation is crucial to the classical approach to econometric modelling, because it implies that if an infinite number of observations is available, the experiment can be repeated an infinite number of times under identical conditions, and consequently the population regression model can be successfully reconstructed. The validity of our results therefore relies decisively on our ability to collect an infinite number of observations. Unfortunately, in a non-experimental discipline like economics, samples are typically small. Given this limitation, the researcher's task of completely recovering the true population regression model seems impossible.

For this reason, there has been a great emphasis by econometricians on the recognition of the value of extraneous information in applied work. Such information provides out of sample knowledge about the population regression model that the researcher is trying to recover. In econometrics, extraneous information can be viewed as imposing certain restrictions on the parameters in the regression model. The sources of such information mainly come from economic theory, or from other empirical studies. If this information is correct, then the combination of such information with that contained in the sample may lead to an increase in estimation efficiency. Extraneous information can be classified into two types : it can either be *a priori* or stochastic. In the former case, the information is precise, while in the latter case, the information is uncertain, has a random component, and is therefore stochastic. In either case, however, the population parameters of interest are still taken as fixed.

There is a separate literature that deals with estimation subject to stochastic extraneous information,[1] and its exploration is beyond the scope of our discussion. In this thesis we shall concentrate on the first type of extraneous information, *i.e.*, the non-sample information is of an *a priori* nature. We will also narrow our attention to cases in which the *a priori* information is in the form of linear restrictions.

Consider a parameter of interest, say, θ. The most general form of linear *a priori* information regarding θ can be represented in the following manner: $r_1 \leq \theta \leq r_2$, where both r_1 and r_2 belong to the Euclidean space. If this information is correct, then the efficiency of our estimator for θ may increase by incorporating this information into the estimation process. The extent to which efficiency will improve depends on how specific our *a priori* information is regarding the actual value of θ. In the above example, the worst scenario occurs when $r_1 = -\infty$ and $r_2 = \infty$ respectively. In this case, we have no specific knowledge about θ, except to say that it is a parameter in the Euclidean space. Consequently, no additional information can be added to that contained in the sample. A one-step improvement from this worst scenario is for either r_1 or r_2

---

[1] See for examples, Theil and Goldberger (1961), Theil (1963) and Brook and Wallace (1973).

to be a finite real number, in which case the true value of θ is known to lie in a half space. Under this scenario, the researcher's knowledge about θ is half way between nothing and complete. Yet a better still scenario occurs if r_1 and r_2 are both known to be finite. In this case, the true value of θ is constrained within an Euclidean subspace with finite bounds. Clearly, the narrower is the interval between r_1 and r_2, the more useful is the incorporation of this prior knowledge in terms of improving an estimator's efficiency. The best scenario occurs when the interval is shrunk to zero, which implies $r_1 = r_2$. In this instance the exact value of θ is known *a priori*, and the sample information plays no part in the process of estimating θ.

In the context of the linear regression model, θ can be thought of as the regression coefficient vector in the model. When there is no specific prior knowledge about the values of the regression coefficents, ordinary least squares (OLS) uses all of the information at hand (in this case sample information only) to generate the best linear unbiased estimator. On the other hand, if some of the exact values of the regression coefficients, or linear combinations of these are known *a priori*, then restricted least squares (RLS), which is a solution obtained by minimizing the sum of squared residuals subject to the exact linear restrictions, can be used to estimate the regression coefficients. The sampling properties of both the OLS and RLS estimators are well documented in the literature and undergraduate textbooks. Relatively less known, however, are the properties of the estimators that result when the researcher is faced with an estimation problem for which the prior information is less complete and the actual values of the regression coefficients are only known to lie in a subspace, which is either bounded from above or below, or both. From a technical view point, these are more complicated problems than the one under the exact restriction scenario, because they involve solving a quadratic programming problem, and the resulting estimators are complicated functions of stochastic variables. Because of the statistical complexities involved, there has been a subsequent lag in the examination of the statistical properties of the estimators that take into the account this incomplete prior information. To reduce the degree of complexity, much of the early work, undertaken in the 1960's and 1970's, assumes that there is only one constraint imposed on a regression coefficient, which is either truncated from above or below, *i.e.*, an inequality restriction. The resulting estimator, which takes into account the inequality *a priori* restriction, is commonly known as the inequality restricted estimator (IRE). It was not until the 1980's that exact finite sample results regarding the properties of the inequality estimator with more than one restriction began to gradually emerge. These studies have also laid the foundation for the analysis of estimation subject to interval restrictions that has taken place in recent years (see, for example, Klemn and Sposito (1980), Escobar and Skarpness (1986), Ohtani (1987) and Hasegawa (1991)).

There is a separate literature that deals with interval restricted estimation, which we do not attempt to discuss here. In this thesis we will focus on the case in which the linear restriction is truncated from either above or below. However, even if we assume the prior information to be a single inequality restriction on the regression coefficient vector, there are still questions regarding the properties of the IRE that remain unresolved, providing the motivation for some of the investigations in this thesis. For example, all of the known results in the literature thus far have focused on the estimation of the coefficient vector, while in practice, the application of the linear regression model typically also involves an unknown scale parameter.

The estimation of the scale parameter is necessary if further hypothesis tests are to be carried out, and it is also needed to form standard errors. However, the literature is virtually silent on the properties of the estimator for the scale parameter that takes into account the inequality restriction imposed on the regression coefficients. Furthermore, it has often been argued that in applied situations, regression models frequently may be mis-specified due to, say, unobservable or inaccurately measured data, or over-simplification. Given that model mis-specification is the norm in most applied econometric analysis, it is perhaps unrealistic to consider the properties of estimators in the context of a model that satisfies all of the ideal assumptions of the classical linear regression model. Arguably, the results generated with model mis-specification taken into account should be of more practical relevance to applied researchers. Again, the literature has paid only very scant attention to the effects of model mis-specification on the properties of the IRE.

In practice, further complications will arise if the validity of the prior restriction is in doubt. When estimating the parameters of interest, a researcher will normally wish to take account of the possibility that his prior and sample information may be in conflict with each other. Typically, he then carries out a test to check for the validity of the prior information. The subsequent inferential procedure will create a pre-testing distortion and will further complicate estimators' properties. This complication arises because the choice of the estimator is dependent on the outcome of a test, which makes it a random event. Pre-testing is a commonly adopted strategy in most applied econometric analyses. It does not only occur when the prior information regarding certain parameters of interests is in doubt. Often with the absence of any prior belief regarding the correct specification of the model, pre-test strategies are used to decide if a particular regressor should be included in the model based on the outcome of a t-test. In other instances, a pre-test strategy may be used to decide the method of estimation. For example, after running a regression by OLS, if the Durbin-Watson statistic indicates the presence of first order autocorrelation, then the model may be re-estimated using, for example, the Cochrane - Orcutt iterative procedure. Otherwise, the original regression, estimated by OLS, is retained. Again, the choice of the estimation procedure is a random event determined by the outcome of a preliminary test.

In the context of the standard linear regression model, when the *a priori* information regarding the regression coefficient is exact, the inferential complications created by pre-testing for the properties of the resulting estimator are now well known. This estimator is commonly known as the equality pre-test estimator (EPTE), which is a choice between the OLS and RLS estimators, based on the outcome of the pre-test. Recent developments have extended the standard results in the literature regarding the properties of the EPTE to situations where the underlying data generating process is mis-specified in various ways, such as those concerning the regressor matrix in the model, or the stochastic process underlying the model's disturbances. The properties of the EPTE when the disturbance term violates the usual assumption of being normally distributed have also been explored.

The choice of optimal critical values for the pre-test has also been considered under various optimality criteria. Results on the full finite sample distribution of EPTEs have also begun to emerge, and so have results on the moments of the pre-test estimators after several rounds of testing and model re-specification (commonly called multi-stage pre-test estimators). Chapter 2 surveys many of these developments.

When the *a priori* information is in the form of linear inequalities, and when a preliminary test is performed to check for the validity of the restrictions, the resulting estimator is called the inequality pre-test estimator (IPTE). This estimator chooses between the IRE and the OLS estimator, depending on the outcome of the test. The literature on the properties of the inequality pre-test estimator is far less comprehensive than that in the exact restriction case. This may be attributable partly to the absence of knowledge of the properties of the IRE until the 1960's and the 1970's, and also to the technical complexities involved when testing for inequality restrictions. The procedures for testing multiple inequality restrictions, and the statistical properties of various tests that have been proposed, are still very much part of ongoing research. For these reasons, the early work on inequality pre-test estimation was confined to cases in which there is only one inequality constraint imposed on the regression coefficents in the model. The traditional one-sided t-test is then used in testing the validity of the single inequality restriction. To reduce the level of complexity further, the early investigations assume that the disturbance variance is known and consequently the t-statistic reduces to the standard normal statistic. It was not until the mid to late 1980's that results on the properties of the IPTE under more complicated situations began to appear in the literature.

Given the lag in the literature in examining the statistical properties of the inequality restricted estimator, not surprisingly there are still a considerable number of issues relating to inequality pre-test estimation that remain unexplored. For example, there are no results in the literature relating to the properties of the IPTE of the scale parameter which takes account of an inequality restriction imposed on the regression coefficients.[2] The choice of optimal critical values for the pre-test of an inequality constraint is virtually unexplored, as are other problems such as the effects of various forms of model mis-specification on the properties of the IPTE; the exact distribution of the IPTE; and inequality pre-test estimation with non-normal disturbances, to name a few.

To summarise, the reliability of any frequentist based econometric analysis depends crucially on the availability of an infinite number of observations. Given the scarce nature of economic data, the results of such an analysis are necessarily approximations and the finite sample properties may vary substantially from these approximations. To improve the reliability of their results, econometricans frequently combine *a priori* information concerning the parameters of interest in conjunction with sample information in the estimation process. When the prior information concerns the half space in which the parameters of interest are to lie, the sampling properties of the resulting inequality restricted and inequality pre-test estimators are still relatively unknown. There are still many unresolved, yet important, issues regarding the statistical properties of the inequality restricted and inequality pre-test estimators. This provides the motivation for the investigation in this thesis.

1.2 OBJECTIVE AND OUTLINE OF THE THESIS

This thesis is directed towards expanding our knowledge of the properties of inequality restricted and pre-test estimators. Section 1.1 has indicated

---

[2] Though to be fair, this problem was only recently explored for the case of exact linear restrictions.

that much is still to be learnt about these estimators and the scope of this thesis is necessarily more limited. We shall focus only on aspects of the following problems : Inequality restricted and pre-test estimation of the scale parameter; the effects of model mis-specification through the omission of relevant regressors on the properties of the inequality restricted and pre-test estimators, for both the prediction vector and the scale parameter; and the question of choosing the optimal critical value for the pre-test of an inequality restriction in both the properly specified and mis-specified models. These topics are selected because of their direct relevance to applied econometrics, as is evidenced by the thorough and careful investigations given in the literature for the case when the prior information is exact.

Throughout the thesis, we assume that the *a priori* information is in the form of a single linear inequality constraint imposed on the regression coefficients. The assumption that there is only one restriction has the advantage of keeping our results tractable, and is also a convenient starting point for future extensions[3].

With these objectives in mind, this thesis is organised as follows:

Chapter 2 surveys the relevant literature on pre-test estimation for exact linear restrictions in econometrics. After discussing the implications of such a pre-test strategy in the standard linear regression model, we consider the literature relating to pre-testing of exact restrictions in mis-specified

---

[3] As noted by Thomson (1982), the biggest difficulty in examining the multiple constraint case lies in the fact that the correlation between the parameter estimates needs to be taken into account. Furthermore, when σ^2 is unknown, the exact finite sample distributions for the test statistics for testing multiple inequality restrictions are generally unknown (see, Farebrother (1986)). This places additional difficulties on the analysis of the pre-test estimator of the scale parameter.

models. Special attention will be given to mis-specification of the regressor matrix. This will set the scene for the analysis later in this thesis.

Chapter 3 reviews the literature on inequality restricted and pre-test estimation. We concentrate on this problem in the context of a linear model with a single linear inequality restriction on the regression vector. This is directly related to the theme of the thesis, although the literature on inequality restricted and pre-test estimation under multiple restrictions is also reviewed briefly.

The rest of the thesis will present several new results on aspects of inequality restricted and pre-test estimation in econometrics. Chapter 4 considers inequality restricted and pre-test estimation for the regression prediction vector in the context of a linear regression model in which relevant regressors are omitted. The exact risk, under squared error loss, of the inequality restricted and inequality pre-test predictors will be derived, numerically evaluated and their properties will be contrasted with the situation in which there is no mis-specification in the model. Using two commonly adopted criteria in the literature, the choice of an optimal critical value for the pre-test in both the properly specified and the omitted variable models is considered in Chapter 5.

Chapter 6 examines the problem of estimating the regression scale parameter in the linear regression model. We assume that the linear model is properly specified and that the prior information is in the form of a single linear inequality constraint imposed on the regression model. We derive and evaluate the risks, under squared error loss, of several inequality restricted and pre-test estimators of the scale parameter, and compare the results with those reported in the literature for the exact restrictions case.

Chapter 7 extends the analysis of Chapter 6 by allowing for model mis-specification through the omission of relevant regressors. The results obtained are contrasted with those from Chapters 4 and 5, and also with the literature which assumes that the prior information is exact. Building on the results of Chapter 6 and 7, Chapter 8 explores the issue of the choice of an optimal pre-test size when estimating the scale parameter.

Finally, some concluding remarks and discussion of the implications of our results are given in Chapter 9.

1.3 Performance Measure

In gauging the performance of an estimator, often a specific form of loss or risk function is used. In the literature, the two most commonly adopted criteria for evaluating an estimator's performance are risk under squared error loss and matrix mean squared error (MSE) measures. In this thesis, only risk under squared error loss is considered. However, as the two criteria are closely related, we shall discuss both of them here. These two criteria both have considerable appeal and limitations.

Suppose θ is a $k \times 1$ vector of parameters of interest. Let $\hat{\theta}$ be an estimator of θ. The $(k \times k)$ MSE matrix of $\hat{\theta}$ is defined as

$$\mathrm{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'\right] = \mathrm{Bias}(\hat{\theta})\mathrm{Bias}(\hat{\theta})' + \mathrm{Cov}(\hat{\theta}).$$

The MSE matrix can therefore be interpreted as representing the bias-variance trade-off inherent in estimation. Let's suppose there is another estimator $\tilde{\theta}$. $\hat{\theta}$ is considered to be MSE superior to $\tilde{\theta}$ if and only if $\left(\mathrm{MSE}(\tilde{\theta}) - \mathrm{MSE}(\hat{\theta})\right)$ is a positive semi-definite matrix. This implies that the mean squared error of each individual component of $\hat{\theta}$ is no greater than that of the corresponding component of $\tilde{\theta}$.

A closely related, but weaker criterion is to require the risk under squared error loss of $\hat{\theta}$ to be no greater than that of $\tilde{\theta}$. The risk under squared error loss of $\hat{\theta}$ is defined as

$$\rho(\hat{\theta}, \theta) = E\left[(\hat{\theta} - \theta)'(\hat{\theta} - \theta)\right].$$

The risk of $\tilde{\theta}$ is defined analgously. It is obvious that the risks of $\tilde{\theta}$ and $\hat{\theta}$ are the traces of their respective MSE matrices. If θ is a scalar, then the risk of any estimator of θ is simply its MSE.

The risk under a squared error loss criterion is generally weaker than the matrix MSE criterion because if $\hat{\theta}$ is better than $\tilde{\theta}$ in terms of matrix MSE, its superiority is assured in terms of risk under squared error loss, but not *vice versa*. However, as it is a scalar, the risk of an estimator is easy to compute and its use makes numerical evaluations of an estimator's risk possible.

Both of these criteria have considerable appeal. They penalise errors in estimation based on a squared magnitude, regardless of sign, and hence negative errors will not cancel out positive errors. They also take into account the bias-variance trade-off involved in estimation. In addition, they are typically easy to compute.

The major disadvantage regarding the use of these criteria is that they can be unduly restrictive. The squared error loss structure, for example, penalises errors according to the squared distance from the true value. There are situations in which such evaluation may not be appropriate. For instance, Giles and Giles (1991) argue that when estimating the regression scale, one would like to penalise under-estimation more than over-estimation. Accordingly, it may be more appropriate to consider the performance of an estimator based on an asymmetric, rather than a squared error loss function. It is not hard to think of other examples in economics where the use of alternative loss structures in evaluating estimators' performance is more appropriate.[4]

[4] See, for example, Varian (1975), Zellner (1991), or Giles (1992a).

CHAPTER TWO
PRELIMINARY TEST ESTIMATION WITH EXACT RESTRICTIONS IN THE LINEAR MODEL : A REVIEW

2.1 INTRODUCTION

In this chapter we assume that the extraneous information is of an exact *a priori* nature and is related to the regression coefficents in the linear model. For example, when estimating a demand equation, the sum of all of the price and income elasticities is constrained by economic theory to be zero in order to eliminate money illusion. Alternatively, in estimating a log-linear production function, the sum of the regression coefficients (excluding the intercept) is restricted to be unity, if we assume constant returns to scale.

A survey on exact linear restrictions pre-test estimation is incorporated in this thesis prior to the chapter that reviews the background material on inequality restricted and pre-test estimation, for two reasons. First, from the points of view of both estimation and inference, inequality restrictions are more difficult to incorporate than exact restrictions. Much of the ongoing research concerning the sampling properties of inequality restricted and pre-test estimators is based, at least partly, on results from earlier studies that assume that the prior restrictions are exact. Second, as mentioned already in Chapter 1, while problems relating to pre-test estimation with exact linear restrictions have been studied rather extensively in the literature, by comparison many issues concerning the properties of estimators when the prior restrictions are not exact still remain unexplored, providing the motivation for the studies in this thesis. In order for us to assess the extent of this gap and the way in which this thesis fits into the existing literature, it is useful to ascertain the breadth of analyses considered in the exact restrictions case.

The rest of this chapter is organised as follows. We begin in Section 2.2 with an overview of the properties of the unrestricted and exact restricted estimators before moving on to survey the literature that investigates the sampling properties of the exact restrictions pre-test estimator for the coefficient vector in Section 2.3, and for the scale parameter in Section 2.4. We assume that the prior information relates to the regression coefficients. In Section 2.5, we review the literature concerning the question of choosing the optimal critical value for the pre-test. Section 2.6 reviews the growing body of literature that deals with the robustness of pre-test estimators to various forms of model mis-specification. Finally, some concluding remarks appear in Section 2.7.

2.2 SAMPLING PROPERTIES OF THE UNRESTRICTED AND EXACT RESTRICTED ESTIMATORS

Consider the linear model

$$y = X\beta + \varepsilon \qquad , \qquad \varepsilon \sim N(0, \sigma^2 I) \tag{2.1}$$

where y and ε are $n \times 1$; X is $n \times k$, nonstochastic and of rank k; and β is $k \times 1$. In addition to the sample information, suppose there exists uncertain prior information about the coefficient vector β, in the form of a set of linear restrictions given by

$$R\beta = r \qquad , \tag{2.2}$$

where R is $j \times k$, and of rank $j \le k$; j is the number of restrictions; and r is $j \times 1$. Furthermore, R and r are both known.

If the researcher believes that the prior information is inconsistent with the underlying data generating process and ignores the restrictions in the estimation procedure, then it is well known that the unrestricted least squares (or maximum likelihood) estimator of β, which utilizes only sample information, is $\tilde{\beta} = S^{-1}X'y$, where $S = X'X$. Clearly, $\tilde{\beta} \sim N(\beta, \sigma^2 S^{-1})$. As $\tilde{\beta}$ is unbiased, its (matrix) Mean Square Error (MSE) is identical to its covariance matrix, which is $\sigma^2 S^{-1}$. The risk of $\tilde{\beta}$ under squared error loss is given by $\rho(\tilde{\beta}, \beta) = \text{tr}(\text{MSE}(\tilde{\beta})) = \sigma^2 \text{tr}(S^{-1})$. It is also well known that the covariance of $\tilde{\beta}$ coincides with the Cramér-Rao lower bound if ε is Normal. $\tilde{\beta}$ is therefore minimum variance unbiased (MVU) and consequently minimizes the risk under squared error loss among the class of all unbiased estimators, linear or non-linear, in this case.

The unrestricted least squares estimator for σ^2 is $\tilde{\sigma}^2_{LS} = (y - X\tilde{\beta})'(y - X\tilde{\beta})/v = \varepsilon' M \varepsilon / v$, where $v = n - k$ is the degrees of freedom of the model and $M = I - XS^{-1}X'$ is an idempotent, symmetric, matrix of rank v. As $\varepsilon' M \varepsilon/\sigma^2 \sim \chi^2_v$, $E(\tilde{\sigma}^2_{LS}) = \sigma^2$ and $\rho(\tilde{\sigma}^2_{LS}, \sigma^2) = \text{var}(\tilde{\sigma}^2_{LS}) = 2\sigma^4/v$. Although this variance does not attain the Cramér-Rao lower bound, there exists no unbiased estimator of σ^2 with variance smaller than $2\sigma^4/v$. This implies that $\tilde{\sigma}^2_{LS}$ is in fact a minimum variance unbiased estimator. The unrestricted maximum likelihood estimator of σ^2 is $\tilde{\sigma}^2_{ML} = (y - X\tilde{\beta})'(y - X\tilde{\beta})/n$. This estimator has a bias of $-k\sigma^2/n$. However, $\text{var}(\tilde{\sigma}^2_{ML}) = 2\sigma^4 v/n^2$ is smaller than $\text{var}(\tilde{\sigma}^2_{LS})$. Furthermore, $\rho(\tilde{\sigma}^2_{ML}, \sigma^2) = (2v + k^2)\sigma^2/n^2$. If one accepts a trade-off between bias and variance, and aims to look for an estimator with the smallest MSE, then the unrestricted minimum mean squared estimator can be shown to be $\tilde{\sigma}^2_{MM} = (y - X\tilde{\beta})'(y - X\tilde{\beta})/(v+2)$. $\tilde{\sigma}^2_{MM}$ has a bias of $-2\sigma^2/(v+2)$. Furthermore, $\text{var}(\tilde{\sigma}^2_{MM}) = 2\sigma^4 v/(v+2)^2$ and $\rho(\tilde{\sigma}^2_{MM}, \sigma^2) = 2\sigma^4/(v+2)$.

Now suppose the researcher is certain that the prior information given in (2.2) is consistent with the underlying data generating process.

He then incorporates the set of exact linear restrictions given in (2.2) into the estimator, and obtains the restricted estimator $\beta^* = \tilde{\beta} - S^{-1}R'(RS^{-1}R')^{-1}(R\tilde{\beta}-r)$ for the coefficient vector. Note that $\beta^* \sim N\left(\beta + S^{-1}R'(RS^{-1}R')^{-1}\bar{\tau},\ \sigma^2(S^{-1}-S^{-1}R'(RS^{-1}R')^{-1}RS^{-1})\right)$, where $\bar{\tau} = r - R\beta$ is the constraint specification error. $MSE(\beta^*) = \sigma^2\left(I - S^{-1}R'(RS^{-1}R')^{-1}R)S^{-1} + S^{-1}R'(RS^{-1}R')^{-1}\bar{\tau}\bar{\tau}'(RS^{-1}R')^{-1}S^{-1}\right)$, and $\rho(\beta^*,\beta)$ is simply the trace of $MSE(\beta^*)$. The restricted least squares, maximum likelihood and minimum mean squared error estimators of σ^2 are $\sigma^{*2}_{LS} = (y - X\beta^*)'(y - X\beta^*)/(v+j)$, $\sigma^{*2}_{ML} = (y - X\beta^*)'(y - X\beta^*)/n$ and $\sigma^{*2}_{MM} = (y - X\beta^*)'(y - X\beta^*)/(v+j+2)$ respectively. $(y - X\beta^*)'(y - X\beta^*)/\sigma^2 \sim \chi^{2'}_{(v+j;\lambda)}$, where $\lambda = \bar{\tau}'(RS^{-1}R')^{-1}\bar{\tau}/2\sigma^2$ is the non-centrality parameter which measures the validity of the linear restrictions. If the restrictions are correct, then $\bar{\tau} = 0$, $\lambda = 0$ and the distribution of $(y - X\beta^*)'(y - X\beta^*)/\sigma^2$ degenerates to central Chi-square with v+j degrees of freedom. Using the properties of non-central Chi-square random variables, one can show that

$$\rho(\sigma^{*2}_{LS},\ \sigma^2) = 2(2\lambda^2+4\lambda+v+j)\sigma^4/(v+j)^2 \tag{2.3}$$

$$\rho(\sigma^{*2}_{ML},\ \sigma^2) = \left(2(j+v+4\lambda)+(j-k+2\lambda)^2\right)\sigma^4/n^2 \tag{2.4}$$

$$\rho(\sigma^{*2}_{MM},\ \sigma^2) = 2(2\lambda^2+v+j+2)\sigma^4/(v+j+2)^2 \quad . \tag{2.5}$$

Both σ^{*2}_{ML} and σ^{*2}_{MM} are biased estimators. σ^{*2}_{LS} is unbiased and σ^{*2}_{MM} is the minimum mean squared error estimator, if the restrictions are true.

In addressing the question of what estimator to use in practice, we know that although β^* is biased unless $R\beta = r$, regardless of the validity of the restrictions, $\left[cov(\tilde{\beta}) - cov(\beta^*)\right]$ is a positive semi-definite matrix. Consequently, the restricted estimators of the individual coefficients have variances that are no greater than the variances of the unrestricted estimators. Therefore, if the non-sample information is true, β^* should always be chosen in favour of $\tilde{\beta}$, as imposing the true restrictions reduces estimation variance and does not induce any bias. However, when the restrictions are incorrect, the use of β^* causes bias, but at the same time reduces the variance. So, the question of whether to impose the restrictions or not involves a bias-variance trade-off. Toro-Vizcarrondo and Wallace (1968) show that in terms of MSE, β^* is preferred to $\tilde{\beta}$ only if $\lambda \le 1/2$. Wallace (1972) shows that $\rho(\beta^*, \beta)$ is smaller than $\rho(\tilde{\beta}, \beta)$ if $\lambda \le \mu_k tr(S^{-1}R'(RS^{-1}R')^{-1}RS^{-1})/2$, where μ_k is the smallest characteristic root of S. If one is interested in the conditional mean forecasting risk, (*i.e.*, the risk of $X\beta$ rather than the risk of β itself), then this condition reduces[1] to $\lambda \le j/2$. However, λ is unknown in practice, and given that the risk of the restricted estimator will increase without bound as the specification error grows, imposing the restrictions when their validity is unknown can be potentially dangerous.

Similarly, it can be shown that

$$\rho(\sigma^{*2}_{LS}, \sigma^2) \le \rho(\tilde{\sigma}^2_{LS}, \sigma^2); \text{ iff } \lambda \le -1+1/\left(4v^2 + 2jv(v+j)\right)^{1/2}/2v \tag{2.6}$$

$$\rho(\sigma^{*2}_{ML}, \sigma^2) \le \rho(\tilde{\sigma}^2_{ML}, \sigma^2); \text{ iff } \lambda \le k-j-2 + \left[(k-j-2)^2 + j(2k-j-2)\right]^{1/2} \tag{2.7}$$

$$\rho(\sigma^{*2}_{MM}, \sigma^2) \le \rho(\tilde{\sigma}^2_{MM}, \sigma^2); \text{ iff } \lambda \le \left(j(v+j+2)/[2(v+2)]\right)^{1/2} . \tag{2.8}$$

Again, these conditions rely on λ, which is unknown in practice. Furthermore, although the researcher would like to minimize the risks of the estimators of both β and σ^2, conditions (2.6), (2.7) and (2.8) do not coincide with the conditions for the restricted estimator of β to dominate its unrestricted counterpart. This implies that there is always a λ - range over which the desired strategy would be to use a mixture of restricted and unrestricted

---

[1] Notice that the criterion given in Toro-Vizcarrondo and Wallace (1968) is stronger than those derived by Wallace (1972) because if β^* is better than $\tilde{\beta}$ in terms of matrix MSE, its superiority is assured in terms of risk under squared error loss, but not *vice versa*. Accordingly, the two criteria relating to the risk superiority of β^* and $X\beta^*$ are called the first and second weak MSE criteria, whereas the one derived by Toro-Vizcarrondo and Wallace (1968) is called the strong MSE criterion.

estimators to estimate β and σ^2. This suggests that a joint risk function for the estimators of β and σ^2 should be considered so that a single condition for the restricted estimators of β and σ^2 to dominate their respective unrestricted counterparts can be determined. This is still to be explored in the literature and is beyond the scope of this thesis.

Given that λ is unknown, the researcher will have doubt as to whether the proposed restrictions should be incorporated. It is in response to this doubt that he conducts a preliminary test to check for the validity of the prior information.

2.3 PRE-TESTING FOR EXACT LINEAR RESTRICTIONS IN THE LINEAR MULTIPLE REGRESSION MODEL

If the validity of the linear restrictions given in (2.2) is uncertain, a pre-test of the following hypothesis could be performed:[2]

$$H_0:\ R\beta = r \quad vs. \quad H_1:\ R\beta \neq r \qquad . \tag{2.9}$$

This hypothesis is tested typically using the Wald statistic

$$u = \frac{(R\tilde{\beta} - r)'(RS^{-1}R')^{-1}(R\tilde{\beta} - r)/j}{(y - X\tilde{\beta})'(y - X\tilde{\beta})/v} \quad . \tag{2.10}$$

Given the assumptions of the model, u has a non-central F distribution with

---

[2] H_0 is analogous to testing $\lambda = 0$. Given that one might prefer to use the restricted estimator even if $\lambda > 0$, Toro-Vizcarrondo and Wallace (1968) suggest that one should be testing H_0: $\lambda \leq 1/2$ *vs.* H_1: $\lambda > 1/2$ instead, as $\lambda = 1/2$ is the value for which the researcher would switch from the restricted to unrestricted estimator of β according to the strong MSE criterion. Wallace and Toro-Vizcarrondo (1968) provide tables for the critical points of this test. Wallace (1972) suggests that if the weak MSE criterion is used, then the relevant hypothesis is H_0: $\lambda \leq j/2$ *vs.* H_1: $\lambda > j/2$. Tables of critical points for this test are contained in Goodnight and Wallace (1972). However, Bock *et al.* (1973a) argue that it does not matter whether one uses this test or the traditional F test, because the critical points for these tests can always be matched up by varying α from one test to another.

degrees of freedom j and v, and non-centrality parameter λ; *i.e.*, $u \sim F'_{(j,v;\lambda)}$. Under H_0, $\lambda = 0$ and $u \sim F_{(j,v)}$.

Our decision rule for the above test is to reject the null if $u > c$, where c is the critical point of the F statistic with j and v degrees of freedom at the desired type I error level. If the null is rejected we use the unrestricted estimator $\tilde{\beta}$. Alternatively, if we cannot reject H_0, the linear restrictions are believed to be valid and the restricted estimator β^* is used in the estimation process. Accordingly, this two-step procedure results in the following pre-test estimator (PTE):

$$\bar{\beta} = \begin{cases} \tilde{\beta} & \text{if } u \geq c \\ \beta^* & \text{if } u < c \end{cases} = I_{[c,\infty)}(u)\tilde{\beta} + I_{[0,c)}(u)\beta^*, \tag{2.11}$$

where c is the critical value determined by $\int_c^\infty dF_{(j,v)} = \alpha$, for a chosen level of significance, α, and $I_{(.,.)}(u)$ is an indicator function which takes the value of 1 if u falls within the subscripted interval and 0 otherwise.

Bancroft (1944) was the first to examine the impact of preliminary tests of significance on subsequent estimation and inference. One of the examples Bancroft considered was a special case of the above problem. He assumed there are two regressors in the model, all variables are measured as deviations about means and the prior information is a single zero restriction imposed on the second coefficient. In the framework of our model, Bancroft's model is equivalent to $X = (x_1\ x_2)$, $\beta' = (\beta_1\ \beta_2)$, $R = (0\ 1)$ and $r = 0$. The PTE of β_1, denoted as $\bar{\beta}_1$, is a choice between the unrestricted estimator, which includes x_2 in the estimation process, and the restricted estimator, which constrains β_2 to be zero, depending on the outcome of the test. Bancroft shows that $\bar{\beta}_1$ is generally biased. The bias of $\bar{\beta}_1$ depends upon the magnitude of β_2, the degree of collinearity between x_1 and x_2, the critical value (and hence the size) of the pre-test, and the sample size. Other things being equal, an increase in sample size reduces the absolute bias of $\bar{\beta}_1$. Intuitively, this is because the larger the sample size, the more precise is the test and the more likely is the researcher to make the right decision. Toro-Vizcarrondo (1968) extends Bancroft's results by deriving the mean square error (MSE) of $\bar{\beta}_1$, and finds that many of the findings regarding the bias of $\bar{\beta}_1$ given by Bancroft are also valid when considering the MSE of $\bar{\beta}_1$. The exact cumulative distribution function of $\bar{\beta}_1$ is derived by Giles and Srivastava (1993). They also examine the implications of adopting the PTE $\bar{\beta}_1$ in the construction of confidence intervals, and show that the true probability level associated with any confidence interval based on $\bar{\beta}_1$ can exceed the nominal (assumed) probability level if β_2 is close to zero. As $|\beta_2|$ increases from zero, the true confidence level based on $\bar{\beta}_1$ decreases before reaching a minimum at a level below the assumed value, and eventually approaches the assumed confidence level as $|\beta_2| \rightarrow \infty$.

Bancroft's "two regressors, one zero constraint" example was later extended by Larson and Bancroft (1963) to the problem of finding an estimator for the prediction vector when it is not certain if a group of regressors should be included in the estimation procedure. A preliminary F-test is then used to determine if the set of coefficients corresponding to the uncertain regressors are jointly equal to zero. Assuming that the data matrix is orthonormal, Larson and Bancroft derive the bias and risk of the resulting PTE for the prediction vector. They also show that the bias function is unaltered by non-orthogonality.

Both Bancroft's (1944) original example and the extension by Larson and Bancroft (1963) are special cases of the general problem described in (2.1), (2.2) and (2.9). Brook (1972) generalises these studies. He derives the risk of the pre-test estimator $\bar{\beta}$ given in (2.11) under the general restrictions case (see also Bock *et al.* (1973a)). He finds that the risk of $\bar{\beta}$ depends, among other things, on X, the data matrix. This places severe limitations on the generalisation of these findings. Brook shows that the risk of $X\bar{\beta}$, on the other hand, depends on the X matrix only through λ. The results regarding the properties of $X\bar{\beta}$ are therefore more general and tractable. The risks of both $\bar{\beta}$ and $X\bar{\beta}$ can be obtained by considering the following weighted risk function (see Judge and Bock (1978, p. 92)):[3]

$$\begin{aligned} E[(\bar{\beta}-\beta)'W(\bar{\beta}-\beta)] &= \sigma^2 tr(S^{-1}W) - \sigma^2 tr(RS^{-1}WS^{-1}R'(RS^{-1}R')^{-1})p\left[F'_{(j+2,v;\lambda)} \le cj/v\right] \\ &\quad - \bar{\tau}'(RS^{-1}R')^{-1}RS^{-1}WS^{-1}R'(RS^{-1}R')^{-1}\bar{\tau}\left\{p\left[F'_{(j+4,v;\lambda)} \le cj/v\right]\right. \\ &\quad \left. - 2p\left[F'_{(j+2,v;\lambda)} \le cj/v\right]\right\} , \qquad (2.12) \end{aligned}$$

where W denotes a known $k \times k$ weighted matrix. When $W = I$, $E[(\bar{\beta}-\beta)'W(\bar{\beta}-\beta)] = \rho(\bar{\beta}, \beta)$, the risk of $\bar{\beta}$. When $W = X'X$, the weighted risk function becomes

$$\begin{aligned} E[(X\bar{\beta}-X\beta)'(X\bar{\beta}-X\beta)] &= \rho(X\bar{\beta}, E(y)) \\ &= \sigma^2\left\{k + (4\lambda - j)\ p\left[F'_{(j+2,v;\lambda)} \le cj/v\right]\right. \\ &\quad \left. - 2\lambda\ p\left[F'_{(j+4,v;\lambda)} \le cj/v\right]\right\}, \qquad (2.13) \end{aligned}$$

which is the risk of the estimated prediction vector, $X\bar{\beta}$. This risk is equivalent to the risk of $\bar{\beta}$ when the regressors are orthonormal. When expressed in terms of the non-centrality parameter λ, $\rho(X\bar{\beta}, E(y))$ is independent of the data matrix. Brook (1972) compares this risk with the predictive risk functions of $\tilde{\beta}$ and β^*, which are given by $\rho(X\tilde{\beta}, E(y)) = \sigma^2 k$ and $\rho(X\beta^*, E(y)) = \sigma^2(k-j-2\lambda)$ respectively. The risks of $X\tilde{\beta}$, $X\beta^*$ and $X\bar{\beta}$ are shown

[3] Bock *et al.* (1973b) derive the risk of the pre-test estimator implied by the alternative testing procedure suggested by Toro-Vizcarrondo and Wallace (1968). They show that in terms of strong MSE, this pre-test estimator is preferred to the unrestricted estimator if $\lambda \le 1/4$.

in Figure 2.1 as functions of λ (see p. 50). The following features are observed :

i) $\rho(X\tilde{\beta}, E(y))$ is invariant with respect to λ, while $\rho(X\beta^*, E(y))$ is a monotonic increasing function of λ. The risks of $X\tilde{\beta}$ and $X\beta^*$ coincide at $\lambda = j/2$. When $\lambda < j/2$, the restricted predictor is risk superior to the unrestricted predictor. Conversely, when $\lambda > j/2$, the unrestricted predictor is preferred.

ii) When λ is relatively small, $\rho(X\bar{\beta}, E(y))$ lies between $\rho(X\beta^*, E(y))$ and $\rho(X\tilde{\beta}, E(y))$. $\rho(X\bar{\beta}, E(y))$ increases with λ and intersects the risk of the unrestricted estimator in the region $j/4 \le \lambda \le j/2$. The maximum of $\rho(X\bar{\beta}, E(y))$ occurs to the right of the intersection between $\rho(X\bar{\beta}, X\beta)$ and $\rho(X\beta^*, E(y))$. Furthermore, $\rho(X\bar{\beta}, E(y)) \to \rho(X\tilde{\beta}, E(y))$ as $\lambda \to \infty$. Intuitively, the likelihood of rejecting the validity of the linear restrictions is high when λ is very high, which increases the frequency of the unrestricted estimator being chosen as the PTE.

iii) There is always a region in $\lambda \in [0,\infty)$ where $X\bar{\beta}$ has the greatest risk among the three estimators under consideration. That region typically lies between the point where $\rho(X\bar{\beta}, E(y)) = \rho(X\tilde{\beta}, E(y))$ and the point where $\rho(X\bar{\beta}, E(y)) = \rho(X\beta^*, E(y))$. For any given λ, $\rho(X\bar{\beta}, X\beta)$ is always greater than $\min[\rho(X\beta^*, E(y)), \rho(X\tilde{\beta}, E(y))]$. In terms of minimizing the estimator's risk under squared error loss, the best that one can do is to adopt the restrictions when $\lambda \le j/2$, and to ignore the restrictions when $\lambda > j/2$.

iv) $\rho(X\tilde{\beta}, E(y))$ and $\rho(X\beta^*, E(y))$ are equivalent to the risk of $X\bar{\beta}$ when $c = 0$ and $c = \infty$ respectively. Other things being equal, an increase in the value of c will cause $\rho(X\bar{\beta}, E(y))$ to lie closer to $\rho(X\beta^*, E(y))$, as the probability of not rejecting the null will increase as c increases. Conversely, a decrease in c will pull the risk function of $X\bar{\beta}$ closer to that of $X\tilde{\beta}$, as the researcher is more likely to reject the null with a lower critical value. Not surprisingly, as c increases, the maximum difference between $\rho(X\bar{\beta}, X\beta)$ and $\min[\rho(X\beta^*, E(y)), \rho(X\tilde{\beta}, E(y))]$ decreases in the region $\lambda \leq j/2$, but increases in the region $\lambda > j/2$. Furthermore, over the entire critical value space, no single PTE strictly dominates the others.

Given the risk characteristics of $X\bar{\beta}$ and its component estimators, if λ is known, then the strategy for the researcher is clear : never pre-test, and choose between the unrestricted and restricted estimators according to the magnitude of λ. Unfortunately, λ is unknown in practice. Naively imposing the restrictions is not recommended as this estimator's risk can be infinitely large. Consequently, we would like to choose a critical value for the pre-test that is "optimal" in terms of minimizing the estimator's risk. This issue will be addressed in Section 2.5. Further discussion on the properties of $\bar{\beta}$ and $X\bar{\beta}$ can be found in Wallace (1977), Judge and Bock (1978, 1983) and Giles and Giles (1993).

2.4 THE EXACT RESTRICTION PRE-TEST ESTIMATOR OF THE SCALE PARAMETER IN LINEAR REGRESSION

Analogously to the pre-test estimator of β, the estimator of σ^2, after a pre-test of $R\beta = r$, is given by

$$\bar{\sigma}_i^2 = \begin{cases} \tilde{\sigma}_i^2 & \text{if } u \geq c \\ \sigma_i^{*2} & \text{if } u < c \end{cases} = I_{[0,c)}(u)\sigma_i^{*2} + I_{[c,\infty)}(u)\tilde{\sigma}_i^2, \tag{2.14}$$

where i = ML, LS and MM.

Clarke *et al.* (1987a,b) derive and evaluate the risk of $\bar{\sigma}_i^2$, which is shown to be

$$\rho(\bar{\sigma}_i^2, \sigma^2) = 1 + \Big\{4\lambda(n+\delta)^2[\lambda P_{80} + (j+2)P_{60} + vP_{42} - (n+\gamma)P_{40}]$$

$$+ v(v+2)(n+\gamma)^2 - 2(n+\gamma)(n+\delta)[v(n+\gamma) + v(\delta-\gamma)P_{02}$$
$$+ j(n+\delta)P_{20}] + j(n+\delta)^2[2vP_{22} + (j+2)P_{40}]$$
$$+ v(v+2)(\delta-\gamma)(2n+\delta+\gamma)P_{04}\Big\}/((n+\gamma)(n+\delta))^2, \qquad (2.15)$$

where $\delta = \gamma = 0$ if i = ML, $\delta = -k$ and $\gamma = j - k$ if i = LS , $\delta = 2 - k$ and $\gamma = 2 + j - k$ if i = MM and $P_{IL} = P\Big[F'_{(j+I,v+L;\lambda)} \le (cj(v+L))/(v(j+I))\Big]$, I,L =0,1,2...

Figure 2.2 (see p. 50) depicts some typical risk functions of $\bar{\sigma}^2_{ML}$ and its component estimators. The following points may be noted:

(1) When using the ML components, there is always a λ range over which the risk of $\bar{\sigma}^2_{ML}$ exceeds those of both $\tilde{\sigma}^2_{ML}$ and σ^{*2}_{ML}. However, there is no such range where $\bar{\sigma}^2_{ML}$ has smaller risk than both $\tilde{\sigma}^2_{ML}$ and σ^{*2}_{ML} simultaneously. This suggests that when estimating the scale parameter based on the method of maximum likelihood, the results concur qualitatively with those obtained when estimating the prediction vector.

(2) Regardless of the choice of component estimators, an increase in α leads to a more frequent rejection of the null hypothesis and consequently less weight will be given to the unrestricted estimator in (2.14). This has the effect of increasing the minimum risk of the PTE, but at the same time it reduces the maximum of this risk. This suggests the concept of an optimal pre-test size, which we discuss in the next section.

(3) When the LS or MM components are used, the PTE can strictly dominate the corresponding unrestricted estimators for appropriate choices of c (see also Ohtani (1988) and Giles (1991a)).

(4) If a mini-max rule with respect to risk under quadratic loss is adopted, then among the three component estimators considered, the one based on the minimum mean square error rule is preferable when constructing a PTE for σ^2.

As an extension to the work of Clarke *et al.* (1987b), Ohtani (1988) shows that when the critical value of the pre-test is $v/(v+2)$, the PTE based on the minimum mean squared error component is the Stein (1964) estimator. He also shows numerically that there is a family of PTEs which strictly dominate the unrestricted estimator, and postulates that $c = v/(v+2)$ is the critical value for minimizing the risk function of $\bar{\sigma}^2_{MM}$ in this family. This result is proved analytically by Gelfand and Dey (1988) (see also Giles (1990)). Futhermore, when using the least squares component estimators, Giles (1991a) shows that regardless of the value of λ, there always exists a class of PTEs with $c \in (0,1)$, all members of which strictly dominate the unrestricted estimator. The minimum risk member of this class is the PTE with $c = 1$. Giles (1991a) also shows that when $j \leq 2$, the PTE with $c = 1$ also strictly dominates the restricted estimator, suggesting that one should always pre-test even if the restrictions are valid. When $j > 2$, the minimum risk estimator is the restricted estimator for small values of λ, and the PTE with $c = 1$ for larger values of λ. Furthermore, Giles (1990) shows that the first derivative of $\rho(\bar{\sigma}^2_i, \sigma^2)$ with respect to c attains zero when $c = 1$ (for LS), $c = v/(v+2)$ (for MM) and $c = 0$ (for ML) and $c = 0, \infty$ appropriately.

Estimating σ is a somewhat different, but closely related, problem to estimating σ^2. The PTE of σ is constructed in a similar manner to those of β and σ^2 and is given by

$$\bar{\sigma}_i = \begin{cases} \tilde{\sigma}_i & \text{if } u \geq c \\ \sigma^*_i & \text{if } u < c \end{cases} = I_{[0,c)}(u)\sigma^*_i + I_{[c,\infty)}(u)\tilde{\sigma}_i . \tag{2.16}$$

i = ML, LS and MM.

Note that $\sqrt{\bar{\sigma}^2_i}$ is not equal to $\bar{\sigma}_i$. Strictly speaking, it is $\bar{\sigma}_i$, rather than $\sqrt{\bar{\sigma}^2_i}$ that one would use in constructing confidence intervals after pre-tests. Clarke (1990) derives the risk of $\bar{\sigma}_i$ for the ML, LS and MM components under the criterion of squared error loss, and finds that the results are qualitatively similar to those for the PTEs of σ^2.

2.5 THE CHOICE OF OPTIMAL CRITICAL VALUES FOR PRE-TESTS

So far, no dominating estimator exists for any of the problems that we have considered, with the exception of estimating the error variance using the least squares component estimators when the number of restrictions is no greater than 2. Moreover, an increase in the critical value of the pre-test typically reduces the minimum risk of the PTE at the expense of increasing its maximum risk. Given that pre-test strategies are commonly adopted in practice, one would like to choose a critical value which brings the risk of the PTE as close as possible to the minimum risk boundary. The choice of an optimal critical value depends on the pre-test problem being investigated, and on the adopted optimality criterion.

One possibility is to choose the pre-test estimator whose minimum risk is the smallest. In the problems that we have investigated, except when one is estimating the scale parameter using the least squares component and $j \le 2$, such a criterion would always lead to the choice of the restricted estimator (*i.e.*, $c = \infty$). At the other extreme, one might use the mini-max principle and choose the pre-test estimator whose maximum risk is minimized. This generally leads to the trivial solution of $c = 0$, as the maximum risk of any PTE that we have considered is always no smaller than the risk of the unrestricted estimator. The exception to this occurs when one is estimating σ^2 using the LS or MM components, then the mini-max criterion leads to the choice of $c = v/(v+2)$ (for MM), $c = 1$ (for LS), or $c = 0$ as the optimal critical value.[4]

An alternative suggestion, made by Sawa and Hiromatsu (1973), is the

---

[4] Giles (1990, 1991a) shows that $c = v/(v+2)$ and $c = 1$ minimize the risk of $\hat{\sigma}^2_{MM}$ and $\hat{\sigma}^2_{LS}$ respectively (see also Ohtani (1988)). Given the properties of the pre-test estimator, it is clear that these critical values are also the mini-max critical values.

criterion of mini-max regret. Sawa and Hiromatsu[5] consider the problem of estimating the regression coefficients in a model with $j \le n$ zero restrictions and an orthonormal data matrix. As this is equivalent to the problem of estimating E(y), we discuss Sawa and Hiromatsu's results in terms of the mean forecasting risk. Within this framework, Sawa and Hiromatsu (1973) define the regret function $\mathrm{Reg}(\lambda, c)$ of $X\bar{\beta}$ as $\rho(X\bar{\beta}, E(y)) - \inf_c \rho(X\bar{\beta}, E(y))$, where $\inf_c \rho(X\bar{\beta}, E(y))$ is the infimum (which is equal to the minimum in this particular problem) of $\rho(X\bar{\beta}, E(y))$ over the entire critical value space for a particular λ value. Now,

$$\inf_c \rho(X\bar{\beta}, E(y)) = \begin{cases} \rho(X\bar{\beta}, E(y)\,|\,c = \infty) = \rho(X\beta^*, E(y)) & \text{for } \lambda \le j/2 \\ \\ \rho(X\bar{\beta}, E(y)\,|\,c = 0) = \rho(X\tilde{\beta}, E(y)) & \text{for } \lambda > j/2 \end{cases}, \qquad (2.17)$$

and so, "maximum regret" can be thought of as the maximum penality, over the entire range of λ, of choosing to use a $\bar{\beta}$ other than the estimator which minimizes risk. According to this mini-max regret principle, a critical value c^* is considered as optimal if $\sup_\lambda \mathrm{Reg}(\lambda, c^*) \le \sup_\lambda \mathrm{Reg}(\lambda, c)$, for all c. So, to minimize the maximum value of the regret function over all possible values of λ and c, the procedure is to seek c^* such that $\sup_{\lambda \le j/2} \mathrm{Reg}(\lambda, c^*) = \sup_{\lambda > j/2} \mathrm{Reg}(\lambda, c^*)$.

Sawa and Hiromatsu numerically compute the optimal critical value for the special case of $j = 1$. In this case, the F-test for testing (2.9) is equivalent to the t-test. Their results show that for moderate to high degrees of freedom, the optimal critical value for the t-test is nearly constant, lying

---

[5] Gun (1965, 1967) had earlier proposed the use of a mini-max regret function in a similar situation.

between 1.376 and 1.370, and decreasing slightly as the degrees of freedom increase.[6] Building in part on the work of Sawa and Hiromatsu, Brook (1972, 1976) tabulated mini-max regret optimal critical values for the case of multiple restrictions. Brook shows that when the regressors are orthonormal, the optimal critical value is generally close to 2, and increases slightly as the number of restrictions increases, *ceteris paribus*. Brook's results for $j = 1$ match those of Sawa and Hiromatsu's.

An alternative to the mini-max regret criterion is to minimize the average relative risk of the PTE, which is equivalent to minimizing the area between the pre-test risk function and the minimum risk boundary. Assuming that the regressors are orthornormal, Toyoda and Wallace (1976) find that using this criterion, the optimal critical value for estimating the regression coefficients is zero if the number of exact restrictions is less than 5, and is about 2 for a large number of restrictions.

The results of Brook (1976) and Toyoda and Wallace (1976) were later extended by Brook and Fletcher (1981) to cover non-orthonormality of the data matrix. Brook and Fletcher consider the model $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where y is $n \times 1$, X_1 is $n \times (k-j)$, X_2 is $j \times 1$, β_1 and β_2 are conformable with their respective regressors and ε satisfies the usual assumptions. The prior restriction of interest is $\beta_2 = 0$. In their analysis, Brook and Fletcher partition the inverse of the $X'X$ matrix as

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} ,$$

such that the extent of multicollinearity of the data is indicated by t = trace

---

[6] See also Farebrother (1975) for an alternative interpretation of Sawa and Hiromatsu's results.

C_{22}. The higher is the degree of multicollinearity, the larger is the value of t. When the data matrix is orthonormal, $C_{22} = I_j$ and trace(C_{22}) = j. Brook and Fletcher find that when using the criterion of minimizing the average relative risk, the optimal critical value is approximately c_{TW} = v(1+t-4)/(j(v+2)) for t > 4 and 0 otherwise. For a given number of restrictions and degrees of freedom, c_{TW} increases with the value of t, indicating that the restricted estimator will be chosen more frequently as the extent of multicollinearity increases. They also show that when using the mini-max regret criterion, the optimal critical value is approximately $c_B = 1 + t/j$. If the regressors are orthonormal, t = j and the optimal critical values collapse to those previously reported by Toyoda and Wallace (1976) and Brook (1976).

We now turn our attention to the problem of choosing an optimal critical value for the pre-test when estimating the regression scale parameter. When $j \leq 2$ and σ^2 is estimated using the least squares component, Giles (1991a) shows that the PTE corresponding to c = 1 can uniformly dominate all other least squares component pre-test estimators of σ^2. Under this situation the optimal critical value of pre-test is clearly unity regardless of the choice of optimality criteria. The strategy for the researcher is therefore always to pre-test, irrespective of the validity of the restrictions. Apart from this exception, there is no other situation in the pre-test literature that we have considered in which a dominating estimator exists. This leads to the consideration of an optimal critical value using the various criteria suggested in the literature for estimating β.

Using the mini-max regret principle of Brook (1976), Giles and Lieberman (1991) tabulated optimal critical values for the three different estimators of σ^2. They find that regardless of the choice of component estimators, mini-max regret optimal critical values are not invariant to the degrees of freedom and number of restrictions. This constrasts with the findings of Brook (1976) who suggests that certain "rules of thumb" can be applied to choosing an optimal critical value for the pre-test when estimating E(y). Giles and Lieberman also compare the risks of the PTEs based on the mini-max regret optimal critical values with those that use the risk minimizing critical values of 1 (for LS), 0 (for ML) and $v/(v+2)$ (for MM). Although there are exceptions, they find that the use of the minimizing critical values generally leads to smaller risk.[7]

Our discussion in this section has assumed that equal weights are given to all values of λ in $[0,\infty)$. However, both Wallace (1972) and Brook and Fletcher (1981) suggest that in determining the optimal critical value of a pre-test, relatively more weight should be given to smaller values of λ, as the existence of prior information presupposes that the researcher must have at least some belief that the prior information is true. Toyoda and Ohtani (1978) show that if more weight is given to small values of λ, then according to the criterion of minimum average relative risk, the optimal critical values are greater than those derived by Toyoda and Wallace (1976). Further extensions of Toyoda and Ohtani's analysis remains for further research.

2.6 THE EFFECTS OF MODEL MIS-SPECIFICATION ON THE SAMPLING PROPERTIES OF PRE-TEST ESTIMATORS

Our discussion so far is based on the assumption that the regression model used for inference purposes is consistent with the underlying process

---

[7] Though outside the scope of this chapter, it should be noted that Ohtani and Toyoda (1978) have also considered mini-max regret critical values for a pre-test of homogeneity. In the context of the same problem, Toyoda and Wallace (1975) tabulate optimal pre-test sizes, when the objective function is to maximize average efficiency.

generating the observations. Although the assumption of a properly specified regression model has provided a convenient conceptual starting point for much subsequent analyses, it is clear that in most applied situations, econometricians invariably work with models which are mis-specified in one way or another . Such mis-specifications are often due to ignorance, lack of data, or the inability of economic theory to define the correct specification. In its broadest sense, mis-specification covers any mistake in a model's underpinning assumptions, such as errors in specifying the design matrix, or the stochastic assumptions relating to the model's disturbance term.

In this section we will review the literature on the effects of various forms of model mis-specification on the properties of the exact restrictions pre-test estimators. Special emphasis will be given to mis-specification with respect to the model's design matrix, because this is closely related to the theme of Chapter 4, in which we examine the effects of omitting relevant regressors on the properties of the restricted and pre-test estimators when the prior information of interest is in the form of a single linear inequality.

2.6.1 Mis-specification of the design matrix

The most likely and pervasive forms of mis-specifying a model's regressor matrix are those of including irrelevant regressors, excluding relevant regressors or replacing the unobervables with proxy variables. In analysing the effects of such mis-specifications on the properties of the resulting estimators, let us first partition X and β in model (2.1) as

$$\underset{(nxk_1)}{X = [X_1} \mid \underset{(nxk_2)}{X_2]} \qquad \text{and} \qquad \underset{(1xk_1)}{\beta' = [\beta_1'} \mid \underset{(1xk_2)}{\beta_2']}$$

respectively.

If the econometrician mistakenly omits the set of regressors X_2, then the following model is specified :

$$y = X_1\beta_1 + \upsilon, \tag{2.17}$$

where $\upsilon = X_2\beta_2 + \varepsilon$. Alternatively, if a group of k_3 irrelevant regressors is included, then the model used by the econometrican is

$$y = X\beta + Z\gamma + u, \tag{2.18}$$

where Z and γ are n x k_3 and k_3 x 1 respectively. Finally, if measurements on X_2 are unobservable and a set of proxy variables P is used in place of X_2, then the model becomes

$$\begin{aligned} y &= X_1\beta_1 + P\eta + w \qquad (2.19)\\ &= X_p\beta_p + w, \end{aligned}$$

where P is $n \times k_2$, η is $k_2 \times 1$, $w = \varepsilon + X_2\beta_2 - P\eta$, $X_p = [X_1 \mid P]$ and $\beta'_p = [\beta'_1 \mid \eta']$. Notice that (2.19) collaspes to (2.17) if $\eta = 0$, and is structurally identical to (2.18) if $\beta_2 = 0$ and $k_2 = k_3$.

Ohtani (1983) first analysed the effects of mis-specifying the model's regressor matrix on the properties of the pre-test predictor. He examines the statistical properties of the pre-test predictor when proxy variables are included. Ohtani assumes that the hypothesis of interest is $R\beta_p = r$, where R and r are defined earlier. The usual test statistic is given by

$$u_p^* = \frac{(R\tilde{\beta}_p - r)'(RS_p^{-1}R')^{-1}(R\tilde{\beta}_p - r)/j}{(y - X_p\tilde{\beta}_p)'(y - X_p\tilde{\beta}_p)/(n-k)} \quad , \tag{2.20}$$

where $S_p = X'_pX_p$. u_p^* has a doubly non-central F distribution with non-centrality parameters

$$\lambda_1^p = (RS_p^{-1}X'_pX\beta - r)'(RS^{-1}R')^{-1}(RS_p^{-1}X'_pX\beta - r)/2\sigma^2$$

and

$$\lambda_2^p = \beta'X'(I-X_pS_p^{-1}X'_p)X\beta/2\sigma^2.$$

Ohtani shows that the risk of the pre-test predictor is

$$\rho(X\bar{\beta},\ E(y)) = \sigma^2\Big(2\lambda_2 + k + (4\lambda_1-j)P\Big[F''_{(j+2,v;\lambda_1^p,\lambda_2^p)} < c_pj(v+j)/(v(j+2))\Big]$$

$$- 2\lambda_1 P\left[F''_{(j+4,v;\lambda_1^p,\lambda_2^p)} < c_p j(v+j)/(v(j+4))\right]\Big), \qquad (2.21)$$

where c_p is the critical value associated with the pre-test.

In order to evaluate the risk of the pre-test predictor, Ohtani considers a simple model with only two regressors, of which one is unobserved and is replaced by a proxy, and the hypothesis of interest is whether the coefficient that corresponds to the unobervable variable is zero. In the framework of the partitioned model of (2.1), these assumptions imply that $k_1 = k_2 = 1$ and the null hypothesis is $\beta_2 = 0$. As X_2 is unobservable, the test statistic is constructed using the estimate of η in the proxy variable model described in (2.19). Ohtani considers the test statistic $H = \tilde{\eta}^2((X'X)(P'P)-(XP)^2)/(X'X)\tilde{\sigma}_p^2$, where $\tilde{\eta}$ and $\tilde{\sigma}_p^2$ are respectively the unrestricted estimators of η and of the scale parameter in (2.19). If the null hypothesis is rejected, the proxy variable model will be used as a basis of estimation; otherwise, the model will be estimated with the proxy variable deleted. Therefore, the pre-test predictor is given by

$$\bar{y} = \begin{cases} y^* = X_1\beta_1^* & \text{if the null is rejected} \\ \tilde{y} = X_p\tilde{\beta}_p & \text{if the null is not rejected} \end{cases}, \qquad (2.22)$$

where β_1^* is the estimator of β_1 with η constrained to zero and $\tilde{\beta}_p$ is the unrestricted estimator of β_p in (2.19).

Within the context of this simple model, Ohtani derives and numerically evaluates the risk of $\bar{y}$ and compares it with the risks of y^* and $\tilde{y}$. He concludes, among other things, that the risk of the pre-test predictor can be smaller than those of both of its components simultaneously over certain regions of the parameter space, a feature not observed when the model is assumed to be properly specified.

As (2.17) and (2.18) are special cases of (2.19), not surprisingly the risk function of the pre-test predictor for the proxy variable model also encompasses those for the underfitted and overfitted models. As observed earlier, the proxy variable model is structually identical to the overfitted model when $\beta_2 = 0$ and $k_2 = k_3$. Without loss of generality, we can also assume that $Z = P$. Correspondingly, λ_1^p and λ_2^p reduce to $\lambda_1^o = (R_1\beta_1 - r)(RS_o^{-1}R')^{-1}(R_1\beta_1 - r)/2\sigma^2$ and $\lambda_2^o = 0$ respectively, where R_1 is a $j \times k_1$ submatrix of R. Consequently, when the linear model is overfitted, the risk function reduces to

$$\rho(X\bar{\beta},\ E(y)) = \sigma^2\Big(k + (4\lambda_1^o - j)P\Big[F''_{(j+2,v;\lambda_1^0,\lambda_2^0)} < c_0 j(v+j)/(v(j+2))\Big] - 2\lambda_1^o P\Big[F''_{(j+4,v;\lambda_1^0,\lambda_2^0)} < c_0 j(v+j)/(v(j+4))\Big]\Big), \qquad (2.23)$$

where c_0 is defined analogously to c_p.

Except for a minor scaling of the first non-centrality parameter, (2.23) is identical to the risk function of $X\bar{\beta}$ when the model is properly specified. In other words, when the model is overfitted, the risk comparisons of the pre-test predictor with its components are maintained as in the properly specified case over the entire region of the parameter space. Accordingly, the results regarding the choice of the optimal critical value of the pre-test are also unaffected when irrelevant regressors are included in the model. Unaware of Ohtani's (1983) work, Giles (1986) directly approaches the problem of including irrelevant regressors and obtains results and conclusions identical to those given above.

Conversely, when $\eta = 0$, the proxy variable model collapses to the underfitted model given in (2.17). The null hypothesis reduces to $R_1\beta_1 = r$ and consequently the test statistic becomes

$$u_u^* = \frac{(R\tilde{\beta}_1 - r)'(RS_1^{-1}R')^{-1}(R\tilde{\beta}_1 - r)/j}{(y - X_1\tilde{\beta}_1)'(y - X_1\tilde{\beta}_1)/(n-k)}, \qquad (2.24)$$

where $S_1 = X_1'X_1$ and $\tilde{\beta}_1 = S_1^{-1}X_1'y$. u_u^* has a doubly non-central F distribution with non-centrality parameters

$$\lambda_1^U = (RS_1^{-1}X_1'X_2\beta_2 - \bar{\tau}_1)'(RS_1^{-1}R')^{-1}(RS_1^{-1}X_1'X_2\beta_2 - \bar{\tau}_1)/2\sigma^2$$

and

$$\lambda_2^U = \beta_2'X_2'(I-X_1S_1^{-1}X_1')X_2\beta_2/2\sigma^2,$$

where $\bar{\tau}_1 = R_1\beta_1 - r$.

Again, unaware of Ohtani's (1983) work, Mittelhammer (1984) deals directly with the problem of underfitting a model. Within the framework of (2.17), he derives the risks of the unrestricted, restricted and pre-test estimators for the prediction vector which are shown to be

$$\rho(X_1\tilde{\beta}_1,\ E(y)) = k_1 + 2\lambda_2^U, \tag{2.25}$$

$$\rho(X_1\beta_1^{**},\ E(y)) = k_1 + 2\lambda_2^U + 2\lambda_1^U - j \tag{2.26}$$

and

$$\rho(X_1\bar{\beta}_1,\ E(y)) = k_1 + 2\lambda_2^U + (4\lambda_1^U - j)P\left[F''_{(j+2,v;\lambda_1^U,\lambda_2^U)} < c_U j(v+j)/(v(j+2))\right] - 2\lambda_1^U P\left[F''_{(j+4,v;\lambda_1^U,\lambda_2^U)} < c_U j(v+j)/(v(j+4))\right], \tag{2.27}$$

where c_U is defined analogously to c_0 and c_p. (2.25) and (2.26) imply that $X\beta^*$ is risk superior to $X\tilde{\beta}$ iff $\lambda_1^U \le j/2$. If there is no specification error in the model, this condition is precisely the one derived by Wallace (1972). However, the trouble is that λ_1^U also depends on $X_2\beta_2$, the degree of model mis-specification. So even if $\bar{\tau}_1 = 0$, λ_1^U may still exceed j/2 because of the effects of the model specification error. This implies that the use of valid prior information does not necessarily lead to a reduction in risk in an underfitted model. Mittelhammer also shows that the risk difference between the unrestricted and pre-test estimators is no longer bounded as the degree of model mis-specification becomes serious.

Giles *et al.* (1992a) investigate the choice of optimal critical values for the model considered by Mittelhammer, using Brook's (1976) mini-max regret criterion. They find that in an underfitted model, the mini-max regret critical values depend on the degrees of freedom. This contrasts with Brook's (1976) results for a properly specified model. For given degrees of freedom and number of restrictions, Giles *et al.* (1992a) show that the optimal critical value declines monotonically as the degree of model mis-specification increases.

Within the framework of Mittelhammer's model, Giles and Clarke (1989) derive and numerically evaluate the risks of the unrestricted, restricted and pre-test estimators of σ^2 based on the maximum likelihood component, which are given by

$$\rho(\tilde{\sigma}^2_{ML},\ \sigma^2) = \sigma^4\Big[2(v_1+4\lambda_2^U) + (2\lambda_2^U-k_1)^2\Big]/n^2 \tag{2.28}$$

$$\rho(\sigma^{*2}_{ML},\ \sigma^2) = \sigma^4\Big[2[j+v_1+4(\lambda_1^U+\lambda_2^U)] + [j-k_1+2(\lambda_1^U+\lambda_2^U)^2\Big]/n^2 \tag{2.29}$$

and

$$\begin{aligned}\rho(\bar{\sigma}^2_{ML},\ \sigma^2) = \rho(\sigma^2_{ML},\ \sigma^2) &+ n^{-2}[2jv_1P''_{22} - \lambda_1^U jnP''_{20} + 4\lambda_1^U v_1 P''_{42} \\ &+ (j(j+2) - 4\lambda_1^U n)P''_{40} + 4(j+2)\lambda_1^U P''_{60} + 4\lambda_1^{U^2} P''_{80} \\ &+ 4\lambda_2^U(jP''_{24} + 2\lambda_n^U P''_{44}))\end{aligned} \tag{2.30}$$

respectively, where $P''_{IL} = P\Big[F''_{(j+I,v+L;\lambda_1^U,\lambda_2^U)} < c_U j(v+j)/(v(j+I))\Big]$.

These risks functions depend on n, k_1 and the non-centrality parameters λ_1^U and λ_2^U. From their numerical results, Giles and Clarke show that in an underfitted model, σ^{*2}_{ML} is risk superior to $\tilde{\sigma}^2_{ML}$ only if

$$2(j+4\lambda_1^U) + (j+2\lambda_1^U)^2 - 2k_1(j+2\lambda_1^U) + 4\lambda_2(j+2\lambda_1^U) < 0. \tag{2.31}$$

Even if the prior information is perfectly correct so that $\bar{\tau} = 0$, λ_1^U is still non-zero unless $\beta_2 = 0$, or X_1 is orthogonal to X_2. Accordingly, condition (2.31) can still be violated even if the prior information is perfectly correct. So the use of valid prior information does not necessarily guarantee a reduction in estimation risk. This is consistent with Mittelhammer's (1984) results in the case of estimating the prediction vector. Giles and Clarke also show that when the degree of mis-specification is serious enough, $\tilde{\sigma}^2_{ML}$ can uniformly dominate both σ^{*2}_{ML} and $\bar{\sigma}^2_{ML}$. This feature is not observed when one is estimating the prediction vector or when the model is properly specified.

Given Giles' (1986) results on the properties of the pre-test predictor when regressors are wrongly included, it is clear that with a simple re-definition of the non-centrality parameter, the results of Clarke *et al.* (1987a, b) on the properties of pre-test estimators of σ^2 when the model is correctly specified will also apply to the case when regressors are wrongly included. The properties of σ^2 in an overfitted model therefore require no further examination. (see Giles (1987) for a discussion)) The properties of the pre-test estimator of σ^2 in a model with proxy variables are unexplored.

2.6.2 Mis-specification of the error distribution

Although in most regression analyses the random disturbances in the model are assumed to be normally distributed, with many economic or financial data, fat-tailed distributions are known to be more appropriate (see, for example, Mandelbröt (1963), Fama (1965) and Judge and Yancey (1986)). One class of distributions which can produce fat tailed distributions is the spherically symmetric family, of which the normal distribution is a special case. Box (1952) shows that the usual Wald test statistic for linear restrictions is distributed as $F_{(j,v)}$ under the null for all members of spherically symmetric disturbances. Thomas (1972) shows that under the alternative, the distribution of the Wald statistic is dependent on the particular type of spherically symmetric process that the disturbances follow. King (1979) shows that the Wald test is uniformly most powerful invariant for all members of spherically symmetric disturbances.

It was not until recently that results on the properties of the exact restrictions PTE with non-normal disturbances began to emerge in the literature. Assuming that the model's disturbances are compound normal, which covers a wide sub-class of the spherically symmetric family, Giles (1991a) derives the risk functions of the PTE of the prediction vector and of the error variance. She also numerically evaluates these risks assuming that the regression's disturbances follow a multivariate Student-t distribution, which is a a special case of the compound normal distribution. Athough Giles concludes that when estimating the prediction vector, most of the results are qualitatively consistent with the conclusions observed for normal errors, Wong and Giles (1992) show that it is possible for the PTE of the prediction vector to dominate both of its components over certain regions of the λ space, a feature not observed when the errors are normally distributed. Wong and Giles also show that when one is estimating the prediction vector and the errors are multivariate Student-t distributed, the optimal critical value using Brook's (1976) mini-max regret principle are not invariant to υ, the degrees of freedom of the multivariate t distribution. However, for a given υ, the mini-max regret critical values are fairly constant with respect to the number of exact restrictions and the degrees of freedom in the model. For $\upsilon \geq 20$, the mini-max regret critical values differ from 2 by at most only 6.7%, suggesting that Brook's rule of thumb optimal critical value prescribed for normal errors can be applied as an approximation for models with multivariate Student-t disturbances when the degrees of freedom associated with the distribution is unknown.

When estimating the model's scale parameter, Giles (1991a) shows that as the model's disturbances depart from normality, the exact restrictions PTE's risk function generally shifts upwards and converges to the risk of the unrestricted estimator at a slower rate than when the disturbances are normally distributed. Giles also shows that when estimating σ^2 using the least squares component estimators, regardless of the values of υ, there is a family of pre-test estimators with $c \in (0,1]$ which strictly dominate the unrestricted estimators, and that the PTE with critical value equal to one has the minimum risk among the members of this family. This implies that ignoring the restrictions is never the optimal strategy. She suggests that for $\upsilon < 15$, all members of the family of pre-test estimators corresponding to $c \in (0,1]$ also strictly dominate the restricted estimator, suggesting that the preferred strategy is always to pre-test when υ is known to be less than 15, even if the restrictions are valid. When $\upsilon > 15$, Giles' results suggest that pre-testing is still the preferred strategy, unless the researcher has strong *a priori* belief regarding the validity of the restrictions.[8]

These results were extended further by Giles (1991b) to the case of an underfitted model. As in the case of a properly specified model, Giles derived the risk expressions for the pre-test estimators of both the prediction vector and the scale parameter. She shows that most of the results described by Mittelhammer (1984) and Giles and Clarke (1989) assuming normal errors carry over to the wider error assumptions. In particular, when the extent of mis-specification associated with omitting relevant regressors is serious, then imposing restrictions even if they are perfectly valid does not guarantee a reduction in an estimator's risk.[9]

---

[8] Giles (1990) extends these results to the estimator of σ^2 when the ML and MM component estimators are used.

[9] The problem of pre-testing for homogeneity with non-normal disturbances has also been investigated (see Metha (1972), Giles (1992b, 1993)).

2.6.3 Mis-specification of the stochastic assumption of the model's disturbance

Frequently, when estimating a linear model, one may incorrectly specify the stochastic process underlying the model's disturbances. For example, although an AR(4) process is often relevant in a model with quarterly data, when testing for linear restrictions using the traditional Wald test, the researcher implicitly assumes that the model's residuals are white noise. Such a mis-specification is likely to affect the power of the test as well as the properties of resulting estimators. The effects of incorrectly assuming a scalar error covariance matrix on the statistical properties of the linear restrictions pre-test estimators are examined by Albertson (1991, 1993). He considers the cases of autoregressive, moving average and heteroscedastic errors, and shows that the effects of mis-specifying the model's error covariance matrix on the sampling performance of pre-test estimators depend to a large extent on the characteristics of the regressor matrix. Regardless of the form of the true error covariance matrix, Albertson shows that the pre-test estimators for both the coefficient vector and the scale parameter can be dominated uniformly by their respective unrestricted counterparts. This is most common when the data are trended and the errors follow a positive AR(1) or MA(1) process. If the data are non-trending or the disturbances follow a negative auto-regressive process, then pre-testing is generally preferred to ignoring the restrictions. When the errors are heteroscedastic and the data are trended, an increase in the degree of heteroscedasticity is likely to increase the region in which the pre-test estimator has risk lower than the unrestricted estimator.

Albertson (1993) also shows that the mini-max regret critical values for the pre-test will vary according to the form of the true stochastic process of the disturbance terms and the characteristics of the regressor matrix.

Therefore, when the model is mis-specified with respect to the covariance structure of its error terms, any attempt to apply a "rule of thumb" optimal critical value, such as that suggested by Brook (1976), will not necessarily lead to an optimal pre-test risk. Albertson (1993) also extends his analysis to models in which relevant regressors are omitted, and shows that the properties of the exact restrictions pre-test estimators will be further distorted according to the magnitude of the specification error resulting from omitting relevant regressors.

Another study that examines the effects of failing to take account of the model's non-scalar error covariance on the properties of the exact restrictions pre-test estimator is that of Giles *et al.* (1992b). They consider the risk properties of the exact restriction pre-test predictor when the error term follows a generalised autoregressive conditional heterosecdastic (GARCH) process. They show that when the GARCH process is sufficiently strong, it is possible for the pre-test predictor to uniformly dominate both of its components. The preferred strategy under such circumstances is to always pre-test. This constrasts with the results that one obtains when the errors are assumed to be white noise.

2.7 CONCLUSIONS

The focus of this chapter is on the properties of pre-test estimators in linear regression, when the hypothesis of interest is represented by a set of exact linear restrictions on the coefficient vector. When one is estimating the prediction vector, although the PTE is never the minimum risk estimator, its risk is bounded over the entire region of the parameter space. When estimating the scale parameter using the least squares rule, with an appropriate choice of test size, the PTE can uniformly dominate the unrestricted estimator for a relatively small number of restrictions.

This shows that pre-testing is not necessarily a bad thing, and in some cases may even be the most advantageous strategy for the researcher. In other instances, when a strictly dominating estimator does not exist, the literature has provided certain prescriptions for applied workers regarding the pre-test critical values that should be chosen in order to minimize the estimator's risk.

Recent studies also show that many of these known properties of pre-test estimators are likely to be distorted when the underlying model is mis-specified. Many of the findings rely on the extent and type of model mis-specification encountered. As they are typically unknown in practice, few prescriptions can be offered to applied researchers as to what strategy should be undertaken in a realistic situation when regression models are invariably mis-specified, except to say that the correct specification is of paramount importance.

While we have focused on a squared error loss function as a measure of estimators' performance throughout the chapter, results on exact restrictions pre-testing based on alternative loss structure have also begun to emerge (see for example, Giles and Giles (1991)). The literature has also given attention to the problem of multi-stage pre-testing. Examples that are of relevance to the theme of this chapter are the work of Shukla (1979) and Özcam and Judge (1991). Detailed discussions of these topics are beyond the scope of this chapter. A comprehensive survey on recent contributions in this area can be found in Giles and Giles (1993a).

In conclusion, the literature provides a great deal of information on the likely consequences of pre-testing for exact linear restrictions on the statistical properties of subsequent estimators. This literature also provides a useful benchmark against which developments in the literature on pre-test estimation with inequality restrictions can be measured. In the next chapter, we will review the literature relating to the problem of estimation subject to linear inequality restrictions in regression.

Appendix 2A

Figure 2.1

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$ and $X\bar{\beta}$

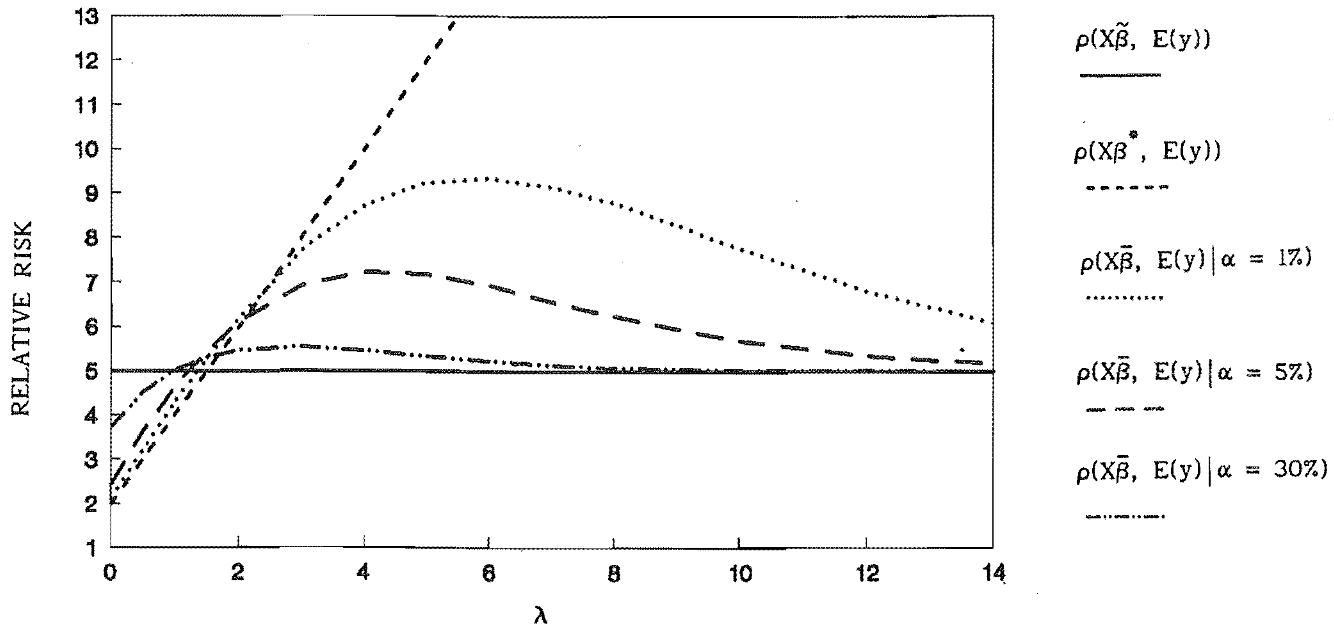
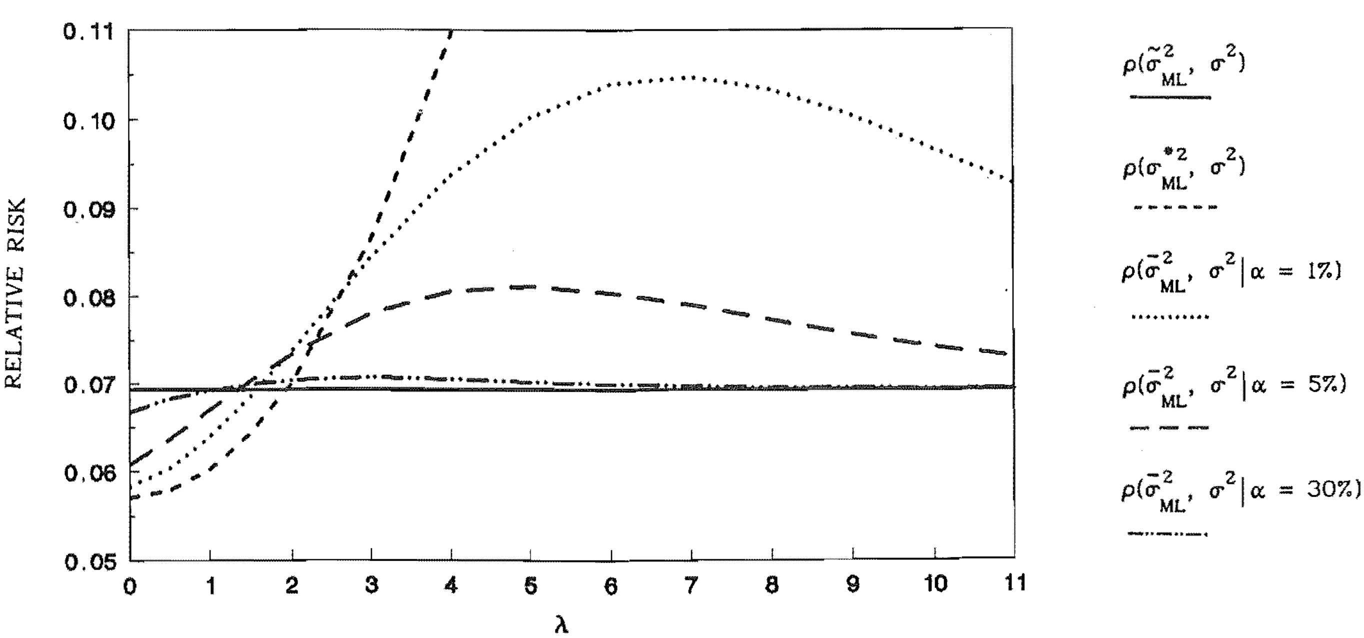


Figure 2.2

Relative risk functions of $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML} and $\bar{\sigma}^2_{ML}$



CHAPTER THREE
INEQUALITY RESTRICTED AND PRE-TEST ESTIMATION IN LINEAR REGRESSION : A SURVEY

3.1 INTRODUCTION

The aim of this chapter is to review the existing literature on the sampling properties of inequality restricted and inequality pre-test estimators in the context of the multiple linear regression model. As mentioned in Chapter 1, our survey will be concerned largely with material relating directly to the theme of this thesis. Accordingly, the rest of this chapter will concentrate mainly on the sampling performance of the inequality restricted and inequality pre-test estimators in a linear regression model with a single linear inequality constraint on the regression coefficients. With only one exception, the literature has focused on the case in which the underlying data generating process is properly specified.

The organisation of the rest of this chapter is as follows. In Section 3.2, we present the assumptions underlying our discussion, introduce the inequality restricted estimator (IRE) and review the literature relating to the standard linear model. The only analysis in the literature on the sampling properties of the IRE in the context of a mis-specified model is the work of Ohtani (1991b), which we review in Section 3.3.

Although not directly related to the theme of this thesis, we will also briefly review the literature on inequality restricted estimation when there are multiple restrictions and illustrate the difference in terms of the sampling properties of the IRE between the one constraint and multiple constraints cases. This will be done in Section 3.4. Section 3.5 considers the pre-test problem for the single inequality constraint case and reviews the literature on the properties of the resulting inequality pre-test estimator (IPTE).

In Section 3.6, we will survey testing procedures, and their properties, for multiple inequality constraints. This is followed by a discussion of the properties of two IPTE for multiple restrictions in Section 3.7. Section 3.8 discusses Ohtani's (1991a) results on the estimation of the error variance after a one sided pre-test of the mean in a normal population. This is closely related to the problem that we consider in Chapter 6. Finally, Section 3.9 offers some concluding remarks, as well as briefly mentioning some related estimators.

3.2 THE STATISTICAL MODEL AND THE INEQUALITY RESTRICTED ESTIMATOR

Consider the following model and prior belief :

$$y = X\beta + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2 I) \tag{3.1}$$

$$C'\beta \geq r \quad , \tag{3.2}$$

where y and ε are $n \times 1$ vectors; X is a full rank $n \times k$ non-stochastic matrix; both β and C are $k \times 1$ vectors and r is a known scalar.

Alternatively, the prior belief may be written as :

$$C'\beta + \bar{\tau} = r. \tag{3.2a}$$

If the direction of the constraint is correct, then $\bar{\tau} \leq 0$. When the restriction holds with strict equality, $\bar{\tau} = 0$.

As maximizing the likelihood function with respect to β is equivalent to minimizing the error sum of squares, the IRE is obtained by solving the following quadratic programming problem :

$$\text{minimize } (y-X\beta)'(y-X\beta) \tag{3.3}$$

subject to

$$C'\beta \geq r \qquad . \tag{3.4}$$

We can form the Lagrangian as

$$L = (y-X\beta)'(y-X\beta) + 2\lambda(r-C'\beta) \quad , \tag{3.5}$$

where λ is the Lagrange multipler.

The Kuhn-Tucker conditions for the above quadratic programming problem are :

$$\partial L/\partial\beta = -2X'y + 2X'X\beta^{**} - 2\lambda^{**}C' = 0, \tag{3.6}$$

$$\partial L/\partial\lambda = r - C'\beta^{**} \leq 0, \tag{3.7}$$

$$\lambda\ (\partial L/\partial\lambda) = \lambda^{**}(r - C'\beta^{**}) = 0, \tag{3.8}$$

and
$$\lambda^{**} \geq 0, \tag{3.9}$$

where λ^{**} and β^{**} represent the solutions to λ and β in the problem.

One can easily observe from (3.6) that the IRE may be expressed as :

$$\beta^{**} = \tilde{\beta} + S^{-1}C'\lambda^{**}, \tag{3.10}$$

where $\tilde{\beta} = S^{-1}X'y$ is the unrestricted estimator (UE) that ignores the prior information.

There are two feasible solutions to the above quadratic programming problem : either the constraint is non-binding, λ^{**} vanishes and β^{**} reduces to $\tilde{\beta}$; or $\lambda^{**} > 0$, in which case the IRE is bounded by the exact prior restriction $C'\beta = r$ and β^{**} reduces to $\beta^{*} = \tilde{\beta} - S^{-1}C(C'S^{-1}C)^{-1}(C'\tilde{\beta}-r)$, the equality restricted estimator (ERE). The first feasible solution is an interior solution while the second feasible solution is a corner point solution. It is well known that in a minimisation problem, the value of the objective function at an interior feasible solution cannot exceed the value at a boundary point feasible solution. Therefore, the conventional unrestricted maximum likelihood estimator is in fact the optimal solution to this problem, unless it violates (3.7), in which case there is no interior solution to the above quadratic programming problem, and the corner point solution given by β^* is then the only feasible, and therefore the optimal, solution. In practice, when faced with an *a priori* restriction of the form given in (3.4), the researcher would typically first fit the regression by least squares, disregarding (3.4). If the resulting unrestricted estimator does not satisfy (3.4), the researcher would run the regression again with the exact *a priori* information imposed on the coefficient vector. Accordingly, the IRE for β is :

$$\beta^{**} = \begin{cases} \tilde{\beta} & \text{if } C'\tilde{\beta} \geq r \\ \beta^* & \text{if } C'\tilde{\beta} < r \end{cases} = I_{[r,\infty)}(C'\tilde{\beta})\tilde{\beta} + I_{(-\infty,r)}(C'\tilde{\beta})\beta^*, \qquad (3.11)$$

where $I_{(.)}(u)$ is an indicator function which takes the value of 1 if u falls in the subscripted interval and 0 otherwise.

The statistical properties of β^{**} are rather complicated because it is a stochastic combination of two non-independently distributed random variables. Although the exact analytical expression for its distribution is unknown, there has been extensive research into the characteristics of its first two moments. Zellner (1961) first studied the sampling properties of the IRE in the context of a simple linear regression model with a non-negativity constraint on the slope coefficient. Zellner analysed the bias of the IRE for the region in which the non-negativity restriction is true and showed that the bias remains small as long as the true value of the slope coefficient does not lie close to the point of truncation of the restriction.

A summary of Zellner's results is given in Malinvaud (1985). Lovell and Prescott (1970) extend Zellner's analysis to the multiple regression model and prove that the IRE always has a smaller MSE than the UE in the region for which the restriction is true if the random disturbances are normally distributed. They find that the reduction in MSE resulting from using β^{**} is maximized when the restriction is a strict equality. They illustrate, however, that even if the restriction is correct, it is possible for the IRE to have a larger MSE than the conventional unrestricted estimator if the disturbances in the model are not normally distributed.

Thomson and Schmidt (1982), building on the work of Lovell and Prescott, compare the MSE of the IRE with those of the UE and the ERE, with the possibility of the non-negativity constraint being incorrectly specified also taken into account. Their results indicate that, given the normality assumption for the disturbance term, β^{**} has a smaller MSE than $\tilde{\beta}$ as long as the non-negativity constraint is in the neighbourhood of being true. β^{**} is preferred to β^{*} only in the region in which the prior information is correct and the true value of β does not lie near the point of truncation of the inequality constraint.

Although these papers assume that the *a priori* restriction is a single non-negativity constraint (*i.e.*, C is a column vector with unity in one of its rows and 0 elsewhere and r = 0), neither the components of the C vector nor the non-negativity part of this assumption are essential for the results. Judge and Yancey (1981) derive the bias and the exact risk (under squared error loss) of the IRE for the more general case of a single linear inequality restriction of the form (3.2). They show that many of the results given above carry over to this more general situation. As the rest of the thesis will build in part on the work of Judge and Yancey (1981, 1986), here we provide a more thorough discussion of their analysis and results.

For convenience, and without loss of generality, Judge and Yancey (1981, 1986) reparameterize the original model (3.1) into an orthornormal model by introducing an orthogonal matrix Q such that $QS^{-1/2}C(C'S^{-1/2}C)^{-1}C'S^{-1/2}Q' = \begin{bmatrix}1 & 0' \\ 0 & 0\end{bmatrix}$. If we let $h = C'S^{-1/2}Q'$, then we can write $h(h'h)^{-1}h' = \begin{bmatrix}1 & 0' \\ 0 & 0\end{bmatrix}$. Furthermore, $h' = h'h(h'h)^{-1}h' = h'\begin{bmatrix}1 & 0' \\ 0 & 0\end{bmatrix} = [h_1, 0]$, where h_1 is the first element of h. Now, the orthonormal transformation is achieved by noting that the original model can be written as:

$$y = Z\theta + \varepsilon, \tag{3.12}$$

where $\theta = QS^{1/2}\beta$ and $Z = XS^{-1/2}Q'$ is a matrix such that $Z'Z = I_k$. Similarly, assuming that h_1 is positive, the prior belief (3.2) may also be transformed as:

$$\theta_1 \geq r_0 \tag{3.13}$$

$$\text{or} \qquad \theta_1 + \tau = r_0 \quad , \tag{3.14}$$

where θ_1 is the first element of θ, $r_0 = r/h_1$ and τ represents the surplus variable, equal to $\bar{\tau}/h_1$.

Using this framework, Judge and Yancey (1981, 1986) derive the mean of the IRE, which can be expressed, when $\tau \leq 0$, as[1]

$$E(\beta^{**}) = \beta + S^{-1/2}Q'\begin{bmatrix} \sigma P(\chi_2^2 \geq \tau^2/\sigma^2)/\sqrt{2\pi} \quad + \tau P(\chi_1^2 \geq \tau^2)/2 \\ 0 \end{bmatrix} . \tag{3.15}$$

When $\tau \geq 0$,

$$E(\beta^{**}) = \beta + S^{-1/2}Q'\begin{bmatrix} \sigma P(\chi_2^2 \geq \tau^2/\sigma^2)/\sqrt{2\pi} \quad + \tau - \tau P(\chi_1^2 \geq \tau^2)/2 \\ 0 \end{bmatrix} . \tag{3.16}$$

Judge and Yancey (1981, 1986) also derive the weighted risk (under squared error loss) of the IRE, which is defined as follows.

---

[1] The assumption that $h_1 > 0$ does not involve any loss of generality. A negative h_1 merely reverses the direction of the inequality in (3.13), in which case the inequality constraint is correct when $\tau \geq 0$ and *vice versa*. The results stated below will still hold with the sign of τ reversed. h_1 is generally non-zero unless C is a null vector, in which case there is no restriction on the regression coefficients.

When $\tau \le 0$,

$$E\left[(\beta^{**}-\beta)'W(\beta^{**}-\beta)\right] = \sigma^2 tr(S^{-1}W) - (\sigma^2/2)a_1 P(\chi_3^2 \ge \tau/\sigma^2) + (\tau^2/2)a_1 P(\chi_1^2 \ge \tau/\sigma^2), \quad (3.17)$$

or alternatively, when $\tau > 0$,

$$E\left[(\beta^{**}-\beta)'W(\beta^{**}-\beta)\right] = \sigma^2 tr(S^{-1}W) + (\sigma^2/2)a_1 P(\chi_3^2 \ge \tau/\sigma^2) - (\tau^2/2)a_1 P(\chi_1^2 \ge \tau/\sigma^2) + a_1(\tau^2 - \sigma^2). \quad (3.18)$$

W is an arbitrary weighting matrix and a_1 is the (1,1) element of the matrix A $= QS^{-1/2}WS^{-1/2}Q'$. When $W = X'X$, $a_1 = 1$ and the weighted risk function collapses to the conditional mean forecasting risk, or the risk of β^{**} assuming orthonormal regressors. When expressed in terms of τ, the conditional mean forecasting risk is quantitatively independent of the data matrix. A typical risk function for β^{**} is plotted in Figure 3.1 (see p. 76). The risks of the UE and the ERE are also shown in that figure for comparison purposes. From the figure and the analytical results, the following may be observed:

(i) The IRE is biased, and its bias is increasing in[2] τ. When $\tau \to \infty$, the bias of the IRE converges to the bias of the ERE.[3] In the region in which the prior belief is true (i.e. $\tau \le 0$), the bias of β^{**} is maximized when the restriction

---

[2] Making use of the results stated in Appendix 4.B of Judge and Yancey (1986, p. 76), when $\tau \le 0$, $\partial(E(\beta^{**})-\beta)/\partial\tau = \left[P(\chi_1^2 \ge \tau^2/\sigma^2)/2\sigma - 2\tau e^{-\tau^2/2\sigma^2}/\tau^2\sqrt{2\pi} \quad 0_{(k-1)}' \right]$, which is always positive. When $\tau > 0$, $\partial(E(\beta^{**})-\beta)/\partial\tau = \left[(1/\sigma - P(\chi_1^2 \ge \tau^2/\sigma^2)/2\sigma \quad 0_{(k-1)}' \right]$, which is also positive.

[3] Using Theorem 2 of Appendix 4C of Judge and Yancey (1986, p. 77), $\lim_{|\tau|\to\infty} |\tau^i| P(\chi_j^2 \ge \tau^2/\sigma^2) \to 0$, for any $i \ge 0$ and $j > 0$. Hence as $\tau \to \infty$, $E(\beta^{**}) \to \beta + [\tau \quad 0_{(k-1)}]$, the risk of the equality restricted estimator.

holds as a strict equality (*i.e.*, $\tau = 0$).[4] The bias of β^{**} approaches zero as τ approaches $-\infty$.

(ii) If the inequality restriction is valid, the risk of the IRE is no greater than that of the UE. If the direction of the inequality constraint is incorrect, the risk of β^{**} is an increasing function of τ and converges to the risk of β^{*} as $\tau \to \infty$. The IRE is preferred to the ERE only when the direction of the inequality constraint is correct and τ is sufficiently small. When $\tau > 0$, the ERE is superior to the IRE. The biggest risk gain from imposing the inequality restriction, as opposed to ignoring the restriction, occurs when $\tau = 0$, at which point the risk of β^{**} is exactly half way between the risks of $\tilde{\beta}$ and β^{*}.

These results show that the major conclusions regarding the sampling properties of the inequality restricted estimator for the single non-negativity constraint case carry over to the slightly more general case where the prior information is a single linear inequality restriction. This indicates that neither the point of truncation nor the functional form of the linear inequality restriction are essential to the results. However, as we shall see, these results may not hold if, (i) there is more than one constraint involved in the model; or (ii) the underlying data generating process is mis-specified. The properties of the IRE when there are multiple inequality constraints will be the theme of our discussion in Section 3.4. Now we focus on the effects of model mis-specification on the properties of the IRE.

---

[4] This is obvious when the bias of β^{**} is an increasing function of τ.

3.3 INEQUALITY RESTRICTED ESTIMATION IN A MIS-SPECIFIED LINEAR REGRESSION MODEL

The literature on inequality restricted estimation has paid relatively scant attention to the effects of model mis-specification on the properties of the inequality restricted estimator. As reported earlier, some preliminary results on the properties of the inequality restricted estimator when the model's random disturbance is not necessarily normally distributed are given in Lovell and Prescott (1970). To the best of our knowledge, no further extension to this analysis is given in the literature[5].

Apart from the numerical example given in Lovell and Prescott (1970), to our knowledge, Ohtani (1991b) is the only other published result on IR estimation under model mis-specification. He considers a linear regression model in which one of the independent variables is unobserved and is replaced by a proxy. He also assumes that there is a non-negativity constraint imposed on the coefficient corresponding to the unobservable variable. The IRE is then a choice between the unrestricted proxy variable estimator and the equality restricted proxy variable estimator, depending on the location of the estimate given by the unrestricted proxy variable estimator relative to the non-negativity restriction. For expository convenience we will call this inequality restricted estimator the proxy variable inequality restricted estimator (PVIRE). Ohtani analyses the bias and the relative efficiency[6] of the PVIRE for the regression coefficient corresponding to the unobservable

---

[5] Prior to Lovell and Prescott's analysis, Rothenberg (1968) independently constructed an example in which he shows that the MSE of the IRE can exceed that of the UE if there is no restriction on the distribution of the model's disturbance.

[6] The relative efficiency of any estimator b is defined here as the ratio between the MSE of the unrestricted estimator and the MSE of b.

variable. He finds that when the correlation between the unobservable variable and its proxy is high, the PVIRE behaves in a very similar way to the IRE for a properly specified model. When the non-negativity restriction is true or in the neighbourhood of being true, the PVIRE is relatively more efficient than the unrestricted proxy variable estimator. When the restriction is invalid, the loss of efficiency from using the PVIRE decreases as the correlation between the unobservable and the proxy variables decreases, and *vice versa.*

Although Ohtani's analysis has provided an interesting insight into the statistical properties of the inequality restricted estimator when the data generating process involves a proxy variable, it has focused only on the properties of the estimator for the coefficient that corresponds to the unobservable variable. It would be equally interesting to observe the effects of the proxy variable on the properties of the inequality restricted estimator for the coefficients of the *observable variables* in the model. In Chapter 4, we shall investigate the properties of the inequality restricted estimator under a different, but commonly encountered form of model mis-specification, namely the omission of relevant regressors.

3.4 ESTIMATION WITH MULTIPLE INEQUALITY RESTRICTIONS

When the *a priori* information involves more than one inequality constraint, the set of feasible solutions to the quadratic programming problem given in Section 3.2 comprises 2^j unrestricted and restricted estimators, where j represents the number of inequality restrictions in the model. As an example, let us consider the case in which the *a priori* belief is represented by $\bar{R}\beta = \begin{bmatrix} C_1' \\ C_2' \end{bmatrix}\beta \geq l = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}$ where $\bar{R}$ is a $2 \times k$ known matrix, C_1 and C_2 are both $k \times 1$ component vectors and l is a 2×1 vector, with known scalars l_1 and l_2 as its elements. In this two restrictions case, the inequality restricted estimator β^{**} can be expressed in the following form:

$$\beta^{**} = \tilde{\beta} \qquad \text{if } C_1'\tilde{\beta} \ge l_1 \text{ and } C_2'\tilde{\beta} \ge l_2 \quad (3.19)$$

$$= \tilde{\beta} - S^{-1}C_1(C_1'S^{-1}C_1)^{-1}(C_1'\beta - l_1) \qquad \text{if } C_1'\tilde{\beta} < l_1 \text{ and } C_2'\tilde{\beta} \ge l_2 \quad (3.20)$$

$$= \tilde{\beta} - S^{-1}C_2(C_2'S^{-1}C_2)^{-1}(C_2'\beta - l_2) \qquad \text{if } C_1'\tilde{\beta} \ge l_1 \text{ and } C_2'\tilde{\beta} < l_2 \quad (3.21)$$

$$= \tilde{\beta} - S^{-1}R'(RS^{-1}R')^{-1}(R\beta - l)\ . \qquad \text{if } C_1'\tilde{\beta} < l_1 \text{ and } C_2'\tilde{\beta} < l_2 \quad (3.22)$$

In this case, one can obtain the optimal solution to the quadratic programming problem by checking if $\tilde{\beta}$ satisfies the restrictions, and accordingly choosing among the four estimators given above. However, in a higher dimensional problem, this procedure will become tedious and time consuming. Several algorithms, such as that suggested by Judge and Takayama (1966) which makes use of the simplex method for quadratic programming developed by Wolfe (1959); and that suggested by Liew (1976), which is based on the algorithm developed by Dantzig and Cottle (1967), can be used to solve the problem. A survey of the various methodologies for solving this quadratic programming can be found in Gill *et al.* (1981). A closed form expression for the inequality restricted estimator with multiple constraints is given in Escobar and Skarpness (1984).

In terms of the sampling properties of the inequality restricted estimator when there are several constraints involved, Liew (1976) considers multiple inequality restrictions of the form $R\beta \ge l$, where R is a $j \times k$ non-stochastic matrix of rank j ($\le$k) and l is a j component vector. Liew proves that, when all the inequality constraints are unbounded, the IRE is asymptotically identical to the UE and is therefore unbiased, consistent and efficient. When some of the constraints are bounded while the rest are unbounded, the IRE reduces to the ERE in large samples. Using a Monte Carlo study, Liew also investigates the small sample properties of the IRE and shows that under the multiple constraints situation, the IRE can have a larger MSE than the UE even if the restrictions are true.[7]

Liew also specifies a small sample covariance matrix for the inequality retricted estimator. However, as noted by Geweke (1986), the validity of Liew's specification is conditional on knowing which restrictions in the model are binding and which are not, and as in practice, the researcher does not possess such knowledge prior to estimation, Liew's specification of the IRE's covariance structure is not of much practical use.

Thomson (1982) considers a two regressor, two non-negativity constraint case, and shows that the properties of the IRE depend not only on the accuracy of the constraints involved, but also on the correlation between the unrestricted estimates of the coefficients in the model. An inaccurate constraint imposed on one coefficient can adversely affect the estimates of the other coefficients in the model, depending on the degree to which the unrestricted estimates are correlated. If the degree of correlation between the unrestricted estimates is zero, then the IRE for the multiple constraints case that Thomson considers has properties identical to the one-constraint case.

Judge and Yancey (1986) consider a multiple regression model with an orthogonal design matrix and orthogonal inequality restrictions. They show that the risk of the IRE is minimised if the true values of all the coefficients lie at the point of truncation of the equality constraints. However, the risk of the IRE will approach infinity if the specification error of any one constraint is infinitely large. Judge and Yancey (1986) also extend

---

[7] However, it must be borne in mind that Liew's experiment is based only on 100 replications, although the possible superiority of the unrestricted estimator over the inequality restricted estimator for the multiple constraints case was later confirmed analytically by Thomson (1982).

their anlaysis to the general design matrix and non-orthogonal restrictions case, and obtain results similar to those obtained for the orthogonal case.[8]

3.5 THE INEQUALITY PRE-TEST ESTIMATOR FOR THE SINGLE INEQUALITY CONSTRAINT CASE

From the results in Section 3.2, we observe that when the prior information is a single linear inequality constraint and the model is properly specified, the IRE, although being biased, has the advantage of having a smaller risk than the conventional unrestricted estimator over a relatively wide range of the parameter space. This typically occurs when the prior belief is correct or nearly so. This suggests that if the researcher is certain of the validity of the prior information, then incorporating the inequality restriction in the estimation process is perhaps a better strategy than ignoring the restriction. However, if the constraint is incorrectly specified, the risk of β^{**} is an increasing function of τ, the constraint specification error. Given that in practice, the researcher typically does not know the magnitude of the constraint specification error, he will typically test the validity of the inequality restriction before deciding whether the prior information should be used. If the test supports the compatibility of the sample information with the prior belief, then the restriction will be imposed on the regression parameters and the IRE will be used; otherwise it will be ignored and the UE will be chosen.

Before we analyse the properties of the estimator generated from this pre-test strategy, let us return once again to model (3.1), in which the prior

---

[8] Judge and Yancey's results rely heavily on the condition that the matrix of restrictions satisfies a monotone property.

belief is a single linear inequality constraint expressed in the form of $C'\beta \geq r$. As seen earlier, this constraint can be reparameterized as $\theta_1 \geq r_0$. Assuming that the disturbance variance σ^2 is known, the test for $H_0 : \theta_1 \geq r_0$ is the standard normal test and the test statistic is given by :

$$(\tilde{\theta}_1 - r_0)/\sigma = u,$$

where $\tilde{\theta}_1$ is the first element of $\tilde{\theta} = QS^{1/2}\tilde{\beta}$, the UE for θ. Our decision rule is to reject the null if $(\tilde{\theta}_1 - r_0)/\sigma$ is less than c, which is the critical value of the test from the standard normal table. Accordingly, the inequality pre-test estimator (IPTE), which chooses between the UE and IRE, based on the outcome of the above test, can be expressed as:

$$\hat{\beta} = \begin{cases} \tilde{\beta} & \text{if } u \leq c \\ \beta^{**} & \text{if } u > c \end{cases} = I_{(-\infty,c]}(u)\tilde{\beta} + I_{(c,\infty)}(u)\beta^{**}. \tag{3.23}$$

If H_0 is not rejected, the estimator chosen would be β^{**}, which is a choice between the UE and ERE, depending on the location of the estimate given by the unrestricted estimator relative to the inequality restriction. However, when $c \geq 0$, $(\tilde{\theta}_1 - r_0)/\sigma > c$ would imply $\tilde{\theta}_1 \geq r_0$ and therefore $C'\tilde{\beta} \geq r$ in terms of the non-orthogonal structure. Hence the UE is always chosen as the estimator when $c \geq 0$, irrespective of whether the null is accepted or not.

Using this framework and assuming that the disturbance variance is known, Judge and Yancey (1986) derive, for $c \leq 0$, the bias of $\hat{\beta}$, which may be expressed in the following way :

When $\tau \leq 0$ and $c+\tau/\sigma \leq 0$,

$$E(\beta^{**}) = \beta + \sigma S^{-1/2}Q'\begin{bmatrix} P(\chi_2^2 \geq \tau^2/\sigma^2) - P(\chi_2^2 \geq (c+\tau/\sigma)^2 \\ 0 \end{bmatrix}/\sqrt{2\pi}$$

$$+ \tau S^{-1/2}Q'\begin{bmatrix} P(\chi_1^2 \geq \tau^2/\sigma^2) - P(\chi_1^2 \geq (c+\tau/\sigma)^2 \\ 0 \end{bmatrix}/2 \quad . \tag{3.24}$$

When $\tau > 0$ and $c+\tau/\sigma \le 0$,

$$E(\beta^{**}) = \beta + \sigma S^{-1/2}Q' \begin{bmatrix} P(\chi_2^2 \ge \tau^2/\sigma^2) \quad - P(\chi_2^2 \ge (c+\tau/\sigma)^2 \\ 0 \end{bmatrix} /\sqrt{2\pi}$$

$$+ \tau S^{-1/2}Q' \begin{bmatrix} 2 - P(\chi_1^2 \ge \tau^2/\sigma^2) \quad - P(\chi_1^2 \ge (c+\tau/\sigma)^2 \\ 0 \end{bmatrix} /2 \; . \qquad (3.25)$$

When $\tau > 0$ and $c+\tau/\sigma > 0$,

$$E(\beta^{**}) = \beta + \sigma S^{-1/2}Q' \begin{bmatrix} P(\chi_2^2 \ge \tau^2/\sigma^2) \quad - P(\chi_2^2 \ge (c+\tau/\sigma)^2 \\ 0 \end{bmatrix} /\sqrt{2\pi}$$

$$+ \tau S^{-1/2}Q' \begin{bmatrix} P(\chi_1^2 \ge (c+\tau/\sigma)^2 \quad - P(\chi_1^2 \ge \tau^2/\sigma^2) \\ 0 \end{bmatrix} /2 \quad . \qquad (3.26)$$

They also derive the weighted risk function of the inequality pre-test estimator, which is expressed as:[9]

When $\tau \le 0$ and $c+\tau/\sigma \le 0$,

$$E\Big[(\hat\beta-\beta)'W(\hat\beta-\beta)\Big] = \sigma^2 tr(S^{-1}W) + a_1\sigma^2\Big[P(\chi_3^2 \ge (c+\tau/\sigma)) - P(\chi_3^2 \ge \tau/\sigma)\Big]/2$$
$$- a_1\tau^2\Big[P(\chi_1^2 \ge (c+\tau/\sigma)) - P(\chi_1^2 \ge \tau/\sigma)\Big]/2 \quad . \qquad (3.27)$$

When $\tau > 0$ and $c+\tau/\sigma \le 0$,

$$E\Big[(\hat\beta-\beta)'W(\hat\beta-\beta)\Big] = \sigma^2 tr(S^{-1}W) - a_1\sigma^2 + a_1\sigma^2\Big[P(\chi_3^2 \ge (c+\tau/\sigma)) + P(\chi_3^2 \ge \tau/\sigma)\Big]/2$$
$$+ a_1\tau^2 - a_1\tau^2\Big[P(\chi_1^2 \ge (c+\tau/\sigma)) + P(\chi_1^2 \ge \tau/\sigma)\Big]/2 \quad . \qquad (3.28)$$

when $\tau > 0$ and $c+\tau/\sigma > 0$,

$$E\Big[(\hat\beta-\beta)'W(\hat\beta-\beta)\Big] = \sigma^2 tr(S^{-1}W) - a_1\sigma^2\Big[P(\chi_3^2 \ge (c+\tau/\sigma)) - P(\chi_3^2 \ge \tau/\sigma)\Big]/2$$
$$+ a_1\tau^2\Big[P(\chi_1^2 \ge (c+\tau/\sigma)) - P(\chi_1^2 \ge \tau/\sigma)\Big]/2 \quad . \qquad (3.29)$$

---

[9] The case of $\tau \le 0$ and $c+\tau/\sigma > 0$ does not exist as c is assumed to be non-positive throughout the analysis.

For comparison purposes, the risk function of $\hat{\beta}$ is also given on Figure 3.1 (p. 76). From the analytical results and the diagram, we observe the following :

(i) By comparing (3.15) - (3.16) with (3.24) - (3.26), we see that the bias of the IPTE is always no greater than that of the IRE. Using the convergence theorem of Judge and Yancey (1986, p. 77), one can show that as $|\tau| \to \infty$, the bias and the risk of $\hat{\beta}$ become asympototic to the bias and the risk of $\tilde{\beta}$ respectively. Intuitively, this is because when τ is sufficiently large, the likelihood of rejecting the null is high and the UE is chosen more frequently. Alternatively, when τ is sufficently small, the chances of accepting the validity of the inequality restriction is high, and so are the chances of the UE not violating the inequality restriction. Hence the UE is again chosen more frequently.

(ii) The bias and risk of the IPTE depend not only on τ, but also on c, the critical value of the pre-test. The IPTE approaches the unrestricted and the inequality restricted estimator when $c \to -\infty$ and $c \to 0$ respectively. The minimum risk boundary of the IPTE is given by either the risk corresponding to $c = -\infty$ or the risk corresponding to $c = 0$.

(iii) There is no region in the parameter space in which the IPTE has the smallest risk. However, there is always a region in the positive horizon of τ in which it has the highest risk among the estimators under consideration. For $c \in (-\infty, 0)$, when τ increases from $-\infty$, the risk of $\hat{\beta}$ decreases before reaching a minimum to the left of the truncation point of the inequality constraint. $\rho(\hat{\beta}, \beta)$ intersects the risk of $\tilde{\beta}$ in the region $0 < \tau < \tau^*$, where τ^* is the intersection point between the risks of the unrestricted and inequality restricted estimators. The maximum of $\rho(\hat{\beta}, \beta)$ occurs to the right of τ^*. The risk of $\hat{\beta}$ eventually approaches the risk of the unrestricted estimator as $\tau \to$

∞.

These results are obtained by assuming that σ^2 is known. When σ^2 is unknown, the test statistic u has a t distribution with n-k degrees of freedom. Judge and Yancey (1986) also derive the bias and the risk of $\hat{\beta}$ for the σ^2 unknown case. Hasegawa (1989a), among other things, derives and numerically evaluates the risk of an inequality pre-test estimator of a normal mean assuming σ^2 is unknown, and shows that qualitatively, the pattern of results are identical to those obtained when σ^2 is known.

3.6 TESTING FOR MULTIPLE INEQUALITY CONSTRAINTS IN THE LINEAR MODEL

When the *a priori* belief consists of more than one inequality restriction, or a mixture of equality and inequality restrictions, the researcher can either test the hypotheses one at a time, or test the entire set of hypotheses jointly. The pre-test estimator that results would depend on the structure of the test adopted by the researcher. Although econometricans have long been familiar with the problem of jointly testing linear equalities on the coefficient vector, they have only recently become aware of the possibility of jointly testing linear inequalities, or linear equalities with one sided alternative. In their seminal paper, Gourieroux *et al.* (1982) consider the problem of testing

$$\bar{H}_0\text{: } R\beta = r \text{ against } \bar{H}_1\text{: } R\beta \geq r \tag{3.30}$$

in the context of the general linear regression model. This testing structure arises when the coefficients in the model are known to satisfy a set of linear inequality constraints, but it is uncertain as to whether the constraints are jointly binding. The Wald statistic for this problem can be constructed by modifying the usual two-sided Wald statistic to

$$W^{**} = (R\beta^{**}-r)'(R(X'\Omega^{-1}X)^{-1}R')^{-1}(R\beta^{**}-r), \tag{3.31}$$

where we have replaced the unrestricted estimators $\tilde{\beta}$ by the inequality restricted estimators β^{**} to take account of the one-sided nature of the alternative hypothesis and Ω is the variance covariance matrix of the model's random disturbances. The Likelihood Ratio test statistic can be constructed in an analogous manner.

Gourieroux *et al.* (1982) show that under the null, both the Wald and the Likelihood Ratio statistics are distributed asymtotically as a "weighted" sum of χ^2 distributions, as opposed to a limiting χ^2 distribution for the two sided case. The statistic of Rao's efficient score test (or the Lagrange Multipler test), on the other hand, is the same as for the two sided case as it takes no account of the one sided nature of the alternative hypothesis and therefore has the usual limiting χ^2 distribution under the null. The use of the efficient score test is not recommended as one would anticipate a loss in power compared with the Likelihood Ratio and Wald tests, which incorporate the prior information. To reconcile this shortcoming, Gourieroux *et al.* (1982) propose a "Kuhn-Tucker" test which is a variant of the Lagrange Multipler test. They also show that the Kuhn-Tucker statistic is asympototically equivalent to the Wald and Likelihood Ratio test statistics. Farebrother (1986) considers a special case of Gourieroux's problem in the context of the standard linear model and derives the exact finite sample distributions of the Likelihood Ratio, Wald and Kuhn-Tucker test statistics under the null hypothesis for the case in which the restriction matrix is of full row rank.[10]

One of the main difficulties with using these test statistics is that they involve the computation of the inequality restricted estimators, which can be

---

[10] Farebrother (1987) shows that the expressions derived in Farebrother (1986) for the null distributions of the statistics are the same as those derived by Hillier (1986) for the special case of $R = [0, I_j]$.

tedious when the number of constraints is large. On the other hand, the Lagrange Multipler test is easier to implement. This motivates Rogers (1986a) to consider a "modified Lagrange Multipler" test which is easy to implement but also takes account of the one sided nature of the alternative hypothesis. However, like the Wald, Likelihood Ratio and Kuhn - Tucker statistics, Rogers' Modified Lagrange Multipler statistic is also distributed as a weighted sum of Chi-square random variables under the null. The weights associated with the null distributions of these statistics are generally unknown and can be difficult to calculate. As a solution to this problem, Rogers (1988) suggested that one should perhaps replace these tests by their corresponding finite induced tests. For example, in testing $R\beta = r$ against $R\beta \geq r$, instead of testing the hypothesis jointly, Rogers suggests that each restriction should be tested separately.[11]

Alternatively, King and Smith (1986) consider a problem where there are j non-zero constraints, *i.e.* $R = [0, I_j]$ and $r = 0$, and suggest that a solution is to rewrite the linear model as $y = X_1\beta_1 + \alpha X_2\beta_2^* + \varepsilon$, where X_1 is $n \times (k-j)$, β_1 is $(k-j) \times 1$, X_2 is $n \times j$, β_2^* is $j \times 1$ known vector, and to conduct a one-sided t test to test $\alpha = 0$ against $\alpha > 0$. They also compare the power of this test with the traditional F test for a two sided alternative hypothesis and the Likelihood Ratio test. They conclude that among these three tests, their test is generally most powerful.[12] A third solution, due to Kodde and Palm

---

[11] Rogers (1986b) shows that the Wald and Likelihood Ratio test statistics for testing (3.30) are equal to certain "infinite induced" test statistics which are variants of the statistic introduced by Scheffé (1953) for the classical two sided problem. A good discussion of infinite induced and finite induced tests is given in Savin (1984).

[12] Farebrother (1990) points out that this may be attributable to King and Smith's use of a positively correlated regressor in their study.

(1986), is to place bounds on the weights. However, as in the case of the corresponding Durbin and Watson testing procedure, one would still need to have recourse to the full procedure when the bounds procedure is inconclusive.

In other instances, one may have no strong belief that the linear inequality restrictions are valid and consquently may like to test for the validity of the inequality constraints against an unrestricted alternative. In this case the hypotheses of interest may be expressed in the form

$$\tilde{H}_0: R\beta \geq 0 \text{ against } \tilde{H}_1: \beta \in R^2 \qquad (3.32)$$

Yancey *et al.* (1981) consider a special case of this situation. They discuss tests of the null hypothesis that a subset of the parameter vector lies in the positive orthant for the orthonormal regression model. Wolak (1987) generalises their results to the case of an arbitrary design matrix and general constraints and proposes the statistic $u^* = (\beta^{**}-\tilde{\beta})'S(\beta^{**}-\tilde{\beta})/\tilde{\sigma}^2$ for testing $\tilde{H}_0$ against $\tilde{H}_1$. Wolak shows that under the null, u^* is distributed as a weighted sum of Chi-square random variables. This work is further extended by Wolak (1989) to the case of the general linear regression model and the linear simultaneous equations model. However, the biggest difficulty with using these tests remains in the calculation of the weights associated with the null distributions. In fact, Wolak (1989) has suggested that it may even be necessary to use simulation techniques to compute these weights when $j \geq 8$ (see also Farebrother (1986)). Using the concept of duality discussed in Wolak (1988), Farebrother (1990) develops two finite induced tests for testing both (3.30) and (3.32). Power comparisons of these finite induced tests and the multivariate tests remain to be explored.

Further discussion of these testing procedures is beyond the scope of this thesis. A survey of some recent developments in testing for inequality constraints is given by Farebrother (1988).

3.7 INEQUALITY PRE-TEST ESTIMATORS FOR THE MULTIPLE CONSTRAINTS CASE

Yancey *et al.* (1989) derive and numerically evaluate the risks of the IPTE that result from the above two testing situations. Although it is found that neither IPTE is uniformly superior, the one which corresponds to the equality null (*i.e.* $\bar{H}_0$) is superior only in the region of the parameter space in which the restrictions hold as strict equalities. When $R\beta < 0$, the risk of the inequality pre-test estimator corresponding to $\bar{H}_0$ increases with the absolute magnitude of the constraint specification error and eventually approaches infinity. This follows from the fact that the IPTE corresponding to $\bar{H}_0$ is constructed under the assumption that $R\beta \geq 0$. The violation of this assumption will naturally lead to a deterioration of the pre-test estimator's risk performance.

In constrast, the IPTE that corresponds to $\tilde{H}_0$ is superior to the unrestricted estimator when $\tilde{H}_0$ is true (*i.e.*, $R\beta \geq 0$). When $R\beta < 0$, the risk of this pre-test estimator increases, but eventually declines and becomes asympototic to the risk of the unrestricted estimator. These results suggest that from the standpoint of avoiding infinite risk, the test structure involving the equality null should be used only if the researcher is certain of the validity of the inequality restrictions. A risk comparison was made between these inequality pre-test estimators and the positive part Stein estimator in Judge *et al.* (1988) for the three parameter case. They find that the positive part Stein estimator is superior to both IPTE over much of the parameter space, and conclude in favour of the use of the positive part Stein estimator.

3.8 A PRE-TEST ESTIMATOR OF THE ERROR VARIANCE IN A MULTIVARIATE NORMAL DISTRIBUTION

The properties of the IPTE for the scale parameter σ^2 after a pre-test for inequality restrictions on the regression coefficients is unexplored in the literature and is the theme of our discussion in Chapters 5 and 6. However, a related problem is considered by Ohtani (1991a). Ohtani is concerned with the estimation of the variance in a normal population after a one-sided pre-test for the mean. Depending on the outcome of the test, the pre-test estimator that Ohtani considers is a choice between the unrestricted and the equality restricted estimators and is therefore different from the problem that we are investigating in this thesis. Here we briefly review Ohtani's results so that comparisons can be made with our analysis later in Chapter 6.

The testing situation that Ohtani considers is the following: H_0: $\mu = \mu_0$ against H_1: $\mu > \mu_0$, where μ is the mean of a normal distribution. If the null is not rejected, then $\mu = \mu_0$ is imposed. Otherwise, $\bar{x}$, the mean from the random sample x_1, x_2, $x_3 \ldots . x_n$, is chosen as the estimator for μ. The corresponding minimum MSE estimator for the variance σ^2 is $s^{*2} = \sum_{i=1}^{n} (x_i - \bar{x})^2/(v+3)$ under H_0 and $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(v+2)$ under H_1. The pre-test estimator $\hat{\sigma}^{*2}$ for σ^2 is then a choice between s^2 and s^{*2}, depending on the outcome of the test. Ohtani proves that the risk of $\hat{\sigma}^{*2}$ reaches a stationary point at $c = (v/(v+2))^{1/2}$, where c is the standard normal critical value of the test and v is the degrees of freedom . He also proves that pre-test estimators with $c < (v/(v+2))^{1/2}$ are inadmissible, being dominated uniformly by the pre-test estimator with $c = (v/(v+2))^{1/2}$. However, in the neighbourhood of H_0, the risk of pre-test estimators with $c > (v/(v+2))^{1/2}$ is smaller than that of the pre-test estimator with $c = (v/(v+2))^{1/2}$.

As the pre-test estimator $\hat{\sigma}^{*2}$ is constructed under the assumption that $\mu \geq$

μ_0, the sampling performance of $\hat{\sigma}^{*2}$ deteriorates when $\mu < \mu_0$ (*i.e.*, both the null and the alternative are incorrect) and approaches infinity as $\mu \to -\infty$. Ohtani also shows that the minimum risk boundary of $\hat{\sigma}^{*2}$ is given by the risk of $\hat{\sigma}^{*2}|c = \infty$ in the neighbourhood of $\mu = \mu_0$, and by the risk of $\hat{\sigma}^{*2}|c = (v/(v+2))^{1/2}$ in the rest of the parameter space. He also shows that the optimal critical value for the pre-test is $c = (v/(v+2))^{1/2}$ under the criterion of minimizing the maximum risk in the parameter space $\mu \geq \mu_0$. Furthermore, under both the null and the alternative, the pre-test estimator with the optimal critical value dominates the Stein (1964) estimator[13] which is the minimum of σ^{*2} and $\tilde{\sigma}^2$.

3.9 CONCLUSIONS

In this chapter we have surveyed the literature on the statistical properties of the IRE and IPTE in the linear regression model. We have concentrated our attention on the assumption that the prior inequality restriction is in the form of a single linear inequality restriction. By and large, the results provided in the literature suggest that when there is only one restriction involved, the properties of both the IRE and IPTE depend, to a large extent, on the accuracy of the inequality restriction. This is qualitatively similar to the results that we obtain when the prior restriction holds as a strict equality.

There are many other estimators which are closely related to the inequality restricted and pre-test estimators that we have discussed above. As they do not directly pertain to the theme of this thesis, we do not aim to

---

[13] The Stein (1964) estimator can be viewed as the pre-test estimator with critical value $(v/(v+2))^{1/2}$ when the alternative hypothesis is two-sided (see Ohtani (1988), Gelfand and Dey (1988), or Ohtani (1991a) for details).

provide a comprehensive discussion of their properties in this chapter. However, for completeness purposes, it is worthwhile to briefly survey the papers which are devoted to the analysis of the their statistical properties.

Throughout our discussion in this chapter, estimation subject to inequality constraints has been treated with the sampling theoretic (classical) approach. However, there also exist studies in the literature, which treat the problem using a Bayesian approach. O'Hagan (1973) considers the estimation of a quadratic regression curve which is constrained to be convex (*i.e.*, the quadratic coefficient is restricted to be positive) using a Bayesian method. Davis (1978) considers multiple linear inequality restrictions in the linear regression model using a Bayesian approach and derives expressions for the probability that the constraints are binding. A somewhat different Bayesian approach, which allows for non-linear inequality constraints and restricts the prior probability of binding constraint to be zero is introduced by Geweke (1986). The risk of a Bayesian inequality pre-test estimator for estimating the sample mean of a normal distribution is derived by Hasegawa (1989a).

Often prior information may exist in the form of an interval constraint. For instance, when a consumption function is estimated, the coefficient of marginal propensity to consume is required by economic theory to lie between zero and unity. The estimator that takes into account the interval restriction on the regression coefficients is called the interval restricted estimator (INRE). If the interval restriction is tested prior to making the decision as to whether the interval constraint should be imposed, then the resulting estimator is called the interval pre-test estimator (INPTE). A closed form expression for the INRE is given by Klemn and Sposito (1980). The sampling properties of the INRE is considered, by Escobar and Skarpness (1986, 1987) and Ohtani (1987). Ohtani (1991c) extends these analyses to cases in which the error term has a multivariate t distribution. Brook and Srivastava (1991) investigate how multicollinearity would affect the MSE of the INRE. A Bayesian variant of the INRE is introduced by Hasegawa (1989b). He considers the sampling properties of this estimator. Finally, some preliminary results on the MSE of the INPTE assuming that the disturbance variance is known are reported in Hasegawa (1991).

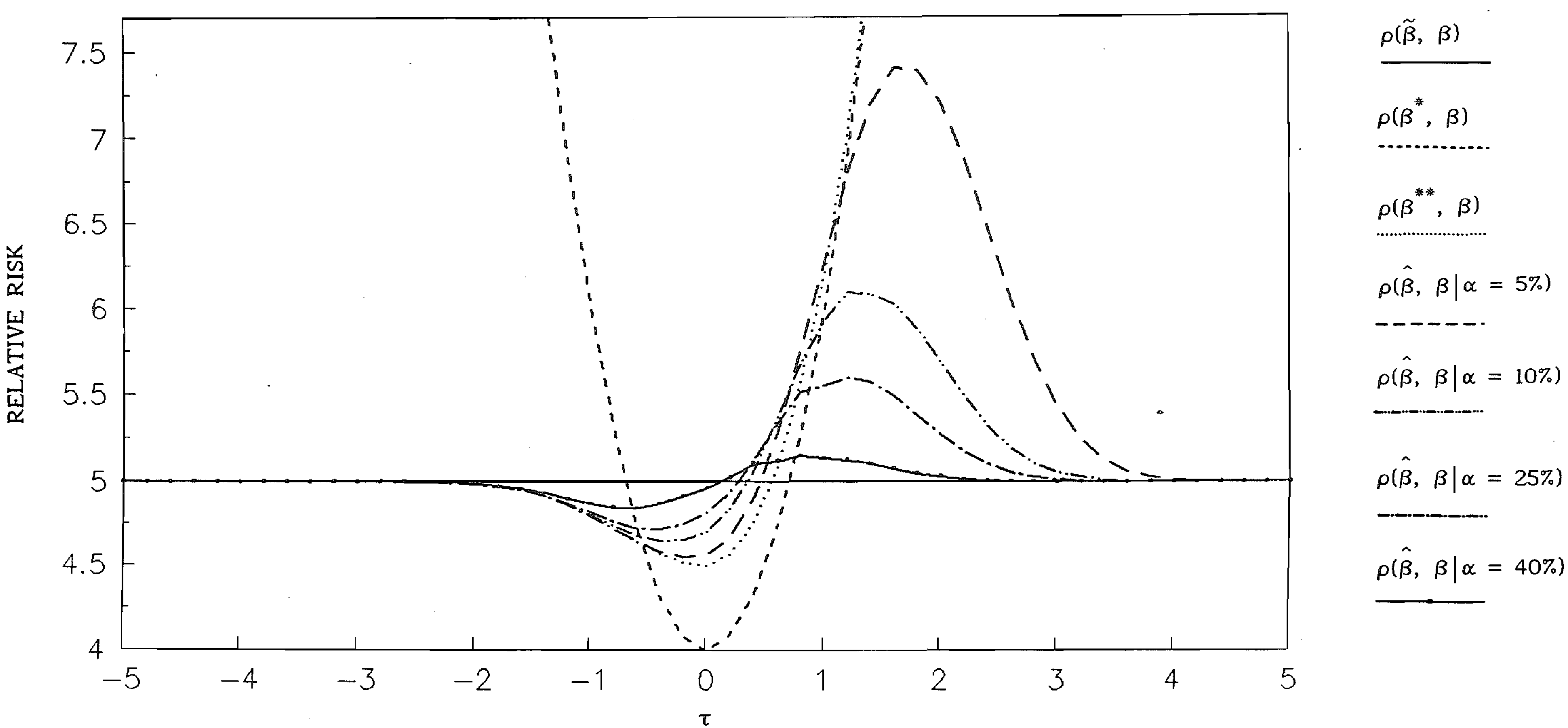
Closely related to the inequality restricted estimator is the family of Stein inequality estimators. Based on an orthonormal linear regression model with a single non-negativity constraint, Judge *et al.* (1984) derive and numerically evaluate the risk functions of the James and Stein inequality restricted estimator and the Stein positive rule inequality restricted estimator. They find that both the traditional inequality restricted estimator and the James and Stein inequality restricted estimator are dominated uniformly by the Stein positive rule inequality restricted estimator. Judge and Yancey (1986) extend this analysis to cases of multiple inequality restrictions.

In the next chapter, we will turn our attention to the properties of the IRE and IPTE estimators when the underlying model is mis-specified through the omission of relevant regressors.

APPENDIX 3A

Figure 3.1

Risk functions of $\tilde{\beta}$, β^*, β^{**} and $\hat{\beta}$



CHAPTER FOUR

THE SAMPLING PROPERTIES OF INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS FOR THE REGRESSION PREDICTION VECTOR UNDER MODEL MIS-SPECIFICATION

4.1 INTRODUCTION

In this chapter we begin our discussion of some new results on the properties of inequality restricted and pre-test estimators that emerge from the research undertaken for this thesis. In the preceding chapter we considered the case of the standard linear regression model, where prior information regarding the regression coefficients is available in the form of an inequality restriction. For this model, when the prior information is correct, the inequality restricted and pre-test estimators are both superior to the traditional estimator that utilises only sample information under a weak mean squared error loss criterion. If the prior information is incorrect, pre-testing is generally superior to naively imposing the restriction. However, with the exception of Ohtani (1991b), who considers inequality restricted estimation in a proxy variable model, the risk comparisons of these estimators have not been examined adequately when the model is mis-specified.

In this chapter, assuming that the prior information is in the form of a single linear inequality restriction on the coefficent vector, we examine the effects of mis-specifying the model through the exclusion of relevant regressors on the sampling performance of the inequality restricted and pre-test estimators for the prediction vector. We consider the estimation of the prediction vector, rather than the coefficient vector itself so that our results are independent of the regressor matrix.[1]

With this objective in mind, this chapter is organised in the following way : for purposes of clarity we first restate the statistical model, the conventional unrestricted and the exact equality restricted predictors in Section 4.2, before going on to derive and evaluate the risk of the inequality restricted predictor in Section 4.3. Section 4.4 considers the testing of an inequality hypothesis in an underfitted model and examines the risk of the corresponding inequality pre-test predictor for the case where σ^2 is known, and also for the case where σ^2 is unknown. Some conclusions that emerge from this investigation are given in the final section.

4.2 THE STATISTICAL MODEL, THE UNRESTRICTED AND EQUALITY RESTRICTED ESTIMATORS

Consider the following data generating process :

$$y = X\beta + Z\eta + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2 I) \tag{4.1}$$

where y and ε are $n \times 1$ vectors; X and Z are non-stochastic matrices of full column rank and are $n \times k$ and $n \times p$ respectively; β and η are unknown coefficient vectors and are $k \times 1$ and $p \times 1$ respectively.

Assume, however, that the model is incorrectly specified as:

$$y = X\beta + \mu \quad . \tag{4.2}$$

So, $\mu = Z\eta + \varepsilon$, and $\mu \sim N(Z\eta, \sigma^2 I)$, but it is assumed by the researcher that $\mu \sim N(0, \sigma^2 I)$. In addition to the sample information, there exists uncertain prior information about the coefficient vector β, in the form of a single linear inequality hypothesis:

---

[1] For completeness, the risks of these estimators for the coefficient vector are derived in Appendix 4B.

$$C'\beta \geq r, \tag{4.3}$$

where C' is a $1 \times k$ known vector and r is a known scalar. Alternatively, one may write (4.3) as

$$C'\beta + \bar{\tau} = r, \tag{4.4}$$

where $\bar{\tau}$ is the surplus variable associated with the inequality restriction. If the direction of the constraint is correct, then $\bar{\tau} \leq 0$. When the restriction holds with strict equality, $\bar{\tau} = 0$.

As mentioned in Chapter 2, the unrestricted estimator (UE) and equality restricted estimator (ERE) of β are $\tilde{\beta} = S^{-1}X'y$ and $\beta^* = \tilde{\beta} - S^{-1}C(C'S^{-1}C)^{-1}(C\tilde{\beta}-r)$ respectively, where $S = X'X$. Throughout this thesis we use risk under quadratic loss as the basis for evaluating the sampling properties of various estimators. Now, if b is any estimator of β in model (4.1), then the risk function of the prediction vector Xb is defined as $\rho(Xb,\ E(y)) = E\left[(Xb - E(y))'(Xb - E(y))\right]/\sigma^2$. Using this definition,[2] Mittelhammer (1984) shows that the predictive risks of $\tilde{\beta}$ and β^* are given by

$$\rho(X\tilde{\beta},\ E(y)) = k + 2\lambda_2 \tag{4.5}$$

and
$$\rho(X\beta^*,\ E(y)) = k + 2\lambda_2 + 2\lambda_1^2 - 1 \tag{4.6}$$

respectively,[3] where $\lambda_1^2 = (\bar{\tau} - C'S^{-1}X'Z\eta)'(C'S^{-1}C)^{-1}(\bar{\tau} - C'S^{-1}X'Z\eta)/(2\sigma^2)$ and $\lambda_2 = \eta'Z'(I-XS^{-1}X')Z\eta/(2\sigma^2)$. The risk of $X\tilde{\beta}$ depends on the model specification

---

[2] We adopt this definition so as to eliminate the nusiance parameter σ^2. This is equivalent to defining $E\left[(Xb - E(y))'(Xb - E(y))\right]$ as the predictive risk function with σ^2 set as 1. This involves no loss of generality.

[3] $(\bar{\tau} - C'S^{-1}X'Z\eta)'(C'S^{-1}C)^{-1}(\bar{\tau} - C'S^{-1}X'Z\eta)/(2\sigma^2)$ is defined as λ_1^2 (rather than λ_1 as commonly adopted in the literature) purely for notational convenience : the risk functions of the inequality restricted and pre-test estimators to be described later are asymmetric in the positive and negative regions of λ_1. Defining λ_1^2 as λ_1 would necessitate the use of the notations $-\sqrt{\lambda_1}$ and $\sqrt{\lambda_1}$ in illustrating these risks.

error through λ_2, while the risk of $X\beta^*$ is conditional on both the constraint and model specification errors through λ_1^2 and λ_2.

Following Judge and Yancey (1986), (4.1) and (4.2) can be reparameterized as :

$$y = H\theta + B\pi + \varepsilon, \tag{4.7}$$

and

$$y = H\theta + \mu \tag{4.8}$$

respectively, where $H = XS^{-1/2}Q'$; $B = ZT^{-1/2}V'$; $\theta = QS^{1/2}\beta$; $\pi = VT^{1/2}\eta$; $T = Z'Z$; $V'V = I_p$, and Q is an orthonormal matrix such that

$$QS^{-1/2}C(C'S^{-1}C)^{-1}C'S^{-1/2}Q' = \begin{pmatrix} 1 & 0' \\ 0 & 0 \end{pmatrix} . \tag{4.9}$$

If we let $h' = C'S^{-1/2}Q'$, then we can rewrite constraint (4.3) as

$$h'\theta \geq r \quad . \tag{4.10}$$

It can be shown, using (4.9), that $h' = (h_1\ 0')$, where h_1 is the first element of the vector h and is assumed to be positive without loss of generality[4]. Using this result, (4.10) can be transformed to

$$\theta_1 \geq r_0 \quad , \tag{4.11}$$

where θ_1 is the first element of θ, and $r_0 = r/h_1$. Alternatively, (4.11) may be written as

$$\theta_1 + \tau = r_0 \quad , \tag{4.12}$$

where $\tau = \bar{\tau}/h_1$ is the surplus variable associated with constraint (4.11).

Now, the unrestricted and equality restricted estimators for θ are :

$$\tilde{\theta} = H'y \tag{4.13}$$

and

$$\theta^* = \tilde{\theta} + h(h'h)^{-1}(r-h'\tilde{\theta})$$

---

[4] See footnote 1 in Chapter 3 for details.

$$= \tilde{\theta} + h(h'h)^{-1}h_1 r_0 - \begin{pmatrix} 1 & 0' \\ 0 & 0 \end{pmatrix}\tilde{\theta}$$

$$= \tilde{\theta} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} r_0 - \begin{pmatrix} \tilde{\theta}_1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} r_0 \\ \tilde{\theta}_{(k-1)} \end{pmatrix} \qquad (4.14)$$

respectively, where $\tilde{\theta}_{(k-1)} = \left(0_1, I_{(k-1)}\right)\tilde{\theta}$.

For the purposes of our analysis, we also reparameterize λ_1^2 and λ_2 as :

$$\begin{aligned} \lambda_1^2 &= (h_1\tau - C'S^{-1/2}Q'QS^{-1/2}X'ZT^{-1/2}V'VT^{1/2}\eta)'(C'S^{-1}C)^{-1} \\ &\quad \times (h_1\tau - C'S^{-1/2}Q'QS^{-1/2}X'ZT^{-1/2}V'VT^{1/2}\eta)/(2\sigma^2) \\ &= (C'S^{-1}Q'(\tau\ 0)' - C'S^{-1/2}Q'H'B\pi)'(C'S^{-1}C)^{-1} \\ &\quad \times (C'S^{-1}Q'(\tau\ 0)' - C'S^{-1/2}Q'H'B\pi)/(2\sigma^2) \\ &= ((\tau\ 0)' - H'B\pi)'\begin{pmatrix} 1 & 0' \\ 0 & 0 \end{pmatrix}((\tau\ 0)' - H'B\pi)/(2\sigma^2) \\ &= ((\tau\ 0)' - H'B\pi)'\begin{pmatrix} 1 \\ 0 \end{pmatrix}(1\ 0)((\tau\ 0)' - H'B\pi)/(2\sigma^2) \\ &= (\tau - \xi)^2/(2\sigma^2), \end{aligned} \qquad (4.15)$$

where $\xi = (H'B\pi)_1$ is the first element of the vector $(H'B\pi)$, and

$$\begin{aligned} \lambda_2 &= \eta'T^{1/2}V'VT^{-1/2}Z'(I - XS^{-1/2}Q'QS^{-1/2}X')ZT^{-1/2}V'VT^{1/2}\eta/(2\sigma^2) \\ &= \pi'B'(I - HH')B\pi/(2\sigma^2) \quad . \end{aligned} \qquad (4.16)$$

4.3 THE INEQUALITY RESTRICTED ESTIMATOR AND ITS PREDICTIVE RISK

According to the two-step procedure described in Chapter 3, the inequality restricted estimator (IRE) in the context of the reparameterized model is

$$\theta^{**} = \begin{cases} \tilde{\theta} & \text{if } \tilde{\theta}_1 \ge r_0 \\ \theta^* & \text{if } \tilde{\theta}_1 < r_0 \end{cases} = I_{(-\infty, r_0)}(\tilde{\theta}_1)\theta^* + I_{[r_0, \infty)}(\tilde{\theta}_1)\tilde{\theta}, \qquad (4.17)$$

where $I_{(.)}(u)$ is an indicator function which takes the value 1 if u falls in the subscripted interval and 0 otherwise.

Recognising that $I_{[r_0,\infty)}(\tilde{\theta}_1) = 1 - I_{(-\infty,r_0)}(\tilde{\theta}_1)$ and $\tau = r_0 - \theta_1$, we can rewrite the IRE as

$$\theta^{**} = \tilde{\theta} - I_{(-\infty,\tau/\sigma)}(u_1)(\tilde{\theta}-\theta^*), \tag{4.18}$$

where $u_1 = (\tilde{\theta}_1-\theta_1)/\sigma$ is a normal random variable with mean ξ/σ and variance 1. Now, if we let $\beta^{**} = S^{-1/2}Q'\theta^{**}$ be the IRE for the coefficient vector in the β space, and using the fact that $\tilde{\theta} - \theta^* = \begin{pmatrix} \tilde{\theta}_1 - r_0 \\ 0_{(k-1)} \end{pmatrix}$, then we can transform (4.18) to

$$\beta^{**} = \tilde{\beta} - S^{-1/2}Q' I_{(-\infty,\tau/\sigma)}(u_1) \begin{pmatrix} \tilde{\theta}_1 - r_0 \\ 0_{(k-1)} \end{pmatrix}$$

$$= \tilde{\beta} - S^{-1/2}Q' \begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0_{(k-1)} \end{bmatrix} \quad . \tag{4.19}$$

When there is no specification error in the model, u_1 collapses to a standard normal random variable, and (4.19) reduces to the expression given in Judge and Yancey (1986, p. 85).

The predictive risk of β^{**} may be written as:

$$\rho(X\beta^{**},\ E(y)) = E\Big[(X\beta^{**}-X\beta-Z\eta)'(X\beta^{**}-X\beta-Z\eta)\Big]/\sigma^2$$

$$= E\Bigg\{\Bigg(X\tilde{\beta} - X\beta - Z\eta - XS^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}\Bigg)'$$

$$.\Bigg(X\tilde{\beta} - X\beta - Z\eta - XS^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}\Bigg)\Bigg\}/\sigma^2$$

$$= \rho(X\tilde{\beta},\ E(y)) + E\Big[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau)^2\Big]/\sigma^2$$

$$-2E\left((X\tilde{\beta} - X\beta - Z\eta)'XS^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}\right)/\sigma^2$$

$$= \rho(X\tilde{\beta},\ E(y)) + E\left[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau)^2\right]/\sigma^2$$

$$-2E\left((\tilde{\theta} - \theta)'QS^{-1/2}X'XS^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}\right)/\sigma^2$$

$$+2\eta'Z'XS^{-1/2}Q'E\begin{bmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}/\sigma^2$$

$$= \rho(X\tilde{\beta},\ E(y)) + E\left[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau)^2\right]/\sigma^2 - 2E\Big[\sigma u_1 I_{(-\infty,\tau/\sigma)}(u_1)$$

$$\times(\sigma u_1-\tau)\Big]/\sigma^2 + 2\xi E\left[I_{(-\infty,\tau/\sigma)}(\sigma u_1-\tau)\right]/\sigma^2.$$

$$= \rho(X\tilde{\beta},\ E(y)) + E\left[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau)(2\xi-\tau-\sigma u_1)\right]/\sigma^2 . \tag{4.20}$$

In evaluating (4.20), we make use of Theorem 4.1, which is a simple generalisation of Theorem 1 given in Judge and Yancey (1986, p. 72) and is stated and proved in Appendix 4A.

Theorem 4.1 forms the basis of the following corollaries :

<u>Corollary 4.1</u>

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)\right] = \begin{cases} \frac{1}{2}P(\chi^2_{(1)} \geq 2\lambda_1^2) & \text{if } \lambda_1 \leq 0 \\ 1-\frac{1}{2}P(\chi^2_{(1)} \geq 2\lambda_1^2) & \text{if } \lambda_1 > 0 \end{cases} . \tag{4.21}$$

<u>Corollary 4.2</u>

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1\right] = \begin{cases} \frac{\xi}{2\sigma}P(\chi^2_{(1)} \geq 2\lambda_1^2) - \frac{1}{\sqrt{2\pi}}P(\chi^2_{(2)} \geq 2\lambda_1^2) & \text{if } \lambda_1 \leq 0 \\ \frac{\xi}{\sigma} - \frac{\xi}{2\sigma}P(\chi^2_{(1)} \geq 2\lambda_1^2) - \frac{1}{\sqrt{2\pi}}P(\chi^2_{(2)} \geq 2\lambda_1^2) & \text{if } \lambda_1 > 0 \end{cases} . \tag{4.22}$$

Corollary 4.3

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^2\right] = \begin{cases} \dfrac{\xi^2}{2\sigma^2}P(\chi^2_{(1)} \geq 2\lambda_1^2) - \dfrac{\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}P(\chi^2_{(2)} \geq 2\lambda_1^2) \\ \qquad + \dfrac{1}{2}P(\chi^2_{(3)} \geq 2\lambda_1^2) \qquad \text{if } \lambda_1 \leq 0 \\ \\ \dfrac{\xi^2}{\sigma^2} - \dfrac{\xi^2}{2\sigma^2}P(\chi^2_{(1)} \geq 2\lambda_1^2) - \dfrac{\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}P(\chi^2_{(2)} \geq 2\lambda_1^2) \\ \qquad + 1 - \dfrac{1}{2}P(\chi^2_{(3)} \geq 2\lambda_1^2) \qquad \text{if } \lambda_1 > 0 \end{cases} \quad . \tag{4.23}$$

When there is no specification error, $\xi = 0$, $\lambda_1 = \tau/(\sqrt{2}\sigma)$ and Corollaries 4.1 to 4.3 collapse to Lemmas 1 to 3 given in Judge and Yancey (1986, Chapter 5).

Making use of Corollaries 4.1 to 4.3 in evaluating[5] (4.19) and recognising that $\lambda_1^2 = (\tau - \xi)^2/(2\sigma^2)$, we can readily show that the risk of $X\beta^{**}$ can be expressed as

$$\rho(X\beta^{**},E(y)) = k + 2\lambda_2 - P_3/2 + \lambda_1^2P_1 \quad , \tag{4.24}$$

where $P_i = P(\chi_i^2 \geq 2\lambda_1^2)$, $i = 1, 3$. Alternatively when $\lambda_1 > 0$, again making use of Corollaries 4.1 to 4.3, the risk of $X\beta^{**}$ can be written as

$$\rho(X\beta^{**},E(y)) = k + 2\lambda_2 - 1 + P_3/2 + 2\lambda_1^2 - \lambda_1^2P_1 \quad . \tag{4.25}$$

These functions are evaluated numerically in order to facilitate their analysis. They are evaluated for $n = 10, 30, 50$, $k = 2, 5$ and $\lambda_1 \in [-10,10]$. The subroutine GAMMQ given in Press *et al.* (1986), is used to calculate P_i, i =

---

5 An alternative way of deriving these risks is to standardize u_1 in (4.18). The argument in the indictor functions of the corresponding risk expressions will then be a standard normal variable, and Lemmas 1 to 3 in Judge and Yancey (1986, Chapter 5) can then be applied directly. The way chosen here has the advantage of pin-pointing how specification error complicates the risk of the estimators.

1, 3. Some representative diagrams are given in Appendix 4C. For comparison purposes, the risk of $X\tilde{\beta}$ and $X\beta^*$ are also included. The case of $\lambda_2 = 0$ is represented in Figure 4.1 (p. 106), which illustrates the results given by Judge and Yancey (1981, 1986). Figures 4.2 and 4.3 (p. 107) illustrate typical cases of $\lambda_2 \neq 0$. We note from these figures that the predictive risk of β^{**}, when $\lambda_2 \neq 0$, depicts essentially the same characteristics as when $\lambda_2 = 0$. In particular, given λ_2, $\rho(X\beta^{**}, E(y))$ is bounded and approaches $\rho(X\tilde{\beta}, E(y))$ as $\lambda_1 \to -\infty$, but it is unbounded and approaches $\rho(X\beta^*, E(y))$ as $\lambda_1 \to \infty$. This result can be verified analytically using the convergence theorem given in Judge and Yancey (1986, p. 77). One can also prove, as shown in Appendix 4A, that for any given λ_2, $\rho(X\beta^{**}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$ when $\lambda_1 \leq 0$. When $\lambda_1 > 0$, the inequality restricted predictor's risk function intersects the risk of the unrestricted predictor and is inferior to the unrestricted predictor's risk over a large portion of the λ_1 space. The biggest risk gain occurs at $\lambda_1 = 0$, where $\rho(X\beta^{**}, E(y))$ is exactly half way between $\rho(X\tilde{\beta}, E(y))$ and $\rho(X\beta^*, E(y))$.

Of course, these results are conditional on λ_1 which varies with ξ, the model specification error, and τ, the surplus variable associated with the inequality restriction. For any given τ, λ_1 can be negative or positive depending on the magnitude of ξ. A non-positive τ, which indicates that the constraint is correctly specified, could result in a positive λ_1, if the model specification error is positive and such that $|\xi| > |\tau|$. It can then be deduced that if ξ is sufficiently large, the predictive risk of β^{**} has the potential to be greater than the unrestricted predictor's risk even when $\tau \leq 0$. Hence the use of valid prior information does not necessarily guarantee a reduction in the estimator's risk. This situation does not arise when the model is correctly specified, in which case both ξ and λ_2 reduce to zero, $\lambda_1 = \tau/(\sqrt{2}\sigma)$ and the inequality $\rho(X\beta^{**}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$ is guaranteed for non-positive τ. This is consistent with Mittelhammer's (1984) results for the case in which the prior information exists in an exact equality form. Finally, given λ_1, $\rho(X\beta^{**}, E(y))$ is unbounded, while $\left[\rho(X\beta^{**}, E(y)) - \rho(X\tilde{\beta}, E(y))\right]$ is bounded by $-P_3/2 + \lambda_1^2 P_1$ (for $\lambda_1 \le 0$), or by $P_3/2 + 2\lambda_1^2 - 1 - \lambda_1^2 P_1$ (for $\lambda_1 > 0$) as $\lambda_2 \to \infty$.

4.4 THE TEST STATISTIC, THE INEQUALITY PRE-TEST PREDICTOR AND ITS RISK

4.4.1. σ^2 known case

Assuming that σ^2 is known, the hypothesis

$$H_0 : \theta_1 \ge r_0 \ \textit{vs.}\ H_1 : \theta_1 < r_0 \tag{4.26}$$

is tested typically using the statistic $\upsilon = (\tilde{\theta}_1 - r_0)/\sigma$. υ has a normal distribution with mean $(\theta_1 + (H'B\pi)_1 - r_0)/\sigma$ and variance 1. Without realising that the model is underfitted, the researcher believes υ to have a standard normal distribution when $\theta_1 = r_0$. Accordingly, the researcher would reject H_0 if $\upsilon \le c$ where c is the size - α critical value for the standard normal variate, and not reject it otherwise. From our discussion in Chapter 3, if the null is rejected, then the UE is chosen, otherwise the IRE is used in the estimation process. Accordingly, the IPTE can be written as:

$$\hat{\theta} = \begin{cases} \tilde{\theta} & \text{if } \upsilon \le c \\ \\ \theta^{**} & \text{if } \upsilon > c \end{cases} = I_{(-\infty,c]}(\upsilon)\tilde{\theta} + I_{(c,\infty)}(\upsilon)\theta^{**} . \tag{4.27}$$

Now, if $c \ge 0$, then accepting H_0 implies $\tilde{\theta} - r_0 > 0$. This condition suggests that the constraint is non-binding and leads to the choice of the unrestricted estimator. Hence when $c \ge 0$, the UE is always chosen regardless of the outcome of the test. Given this result, the properties of the IPTE for the case of $c \ge 0$ needs no further discussion.

In the discussion that follows, we assume that $c < 0$. From our earlier discussion, $\theta^{**} = \tilde{\theta} - \begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix}$ and $I_{(c,\infty)}(\upsilon)$ can be written as 1 $- I_{(-\infty,c]}(\upsilon)$. Using these results, one can easily show that

$$\hat{\theta} = \tilde{\theta} + \begin{pmatrix} I_{(-\infty,c)}(\upsilon)I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) - I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix} . \tag{4.28}$$

$$\begin{aligned} \text{Now, } I_{(-\infty,c)}(\upsilon)I_{(-\infty,\tau/\sigma)}(u_1) &= I_{(-\infty,c)}((\tilde{\theta}_1-\theta_1+\theta_1-r_0)/\sigma)I_{(-\infty,\tau/\sigma)}(u_1) \\ &= I_{(-\infty,c+\tau/\sigma)}(u_1)I_{(-\infty,\tau/\sigma)}(u_1) \\ &= I_{(-\infty,c+\tau/\sigma)}(u_1), \end{aligned} \tag{4.29}$$

as $c < 0$ by assumption.

Therefore,

$$\hat{\theta} = \tilde{\theta} + \begin{pmatrix} [I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)](\sigma u_1-\tau) \\ 0 \end{pmatrix}. \tag{4.30}$$

As $\hat{\theta} = QS^{1/2}\hat{\beta}$, it is straightforward to show that

$$\hat{\beta} = \tilde{\beta} + S^{-1/2}Q'\begin{pmatrix} [I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau)}(u_1)](\sigma u_1-\tau) \\ 0 \end{pmatrix}. \tag{4.31}$$

Using these results, one can readily show that the predictive risk of $\hat{\beta}$ can be expressed as:

$$\begin{aligned} \rho(X\hat{\beta},\ E(y)) = \rho(X\tilde{\beta},\ E(y)) &+ E\left[\left(I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)\right)^2\right. \\ &\left.\times(\sigma u_1-\tau)^2\right]/\sigma^2 + 2E\left[(X\tilde{\beta} - X\beta - Z\eta)'XS^{-1/2}Q'\right. \\ &\left.\times\begin{pmatrix} [I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)](\sigma u_1-\tau) \\ 0 \end{pmatrix}\right]/\sigma^2 \\ = \rho(X\tilde{\beta},\ E(y)) &+ E\left[\left(I_{(-\infty,c+\tau/\sigma)}(u_1) + I_{(-\infty,\tau/\sigma)}(u_1)\right.\right. \\ &\left.\left.- 2I_{(-\infty,\tau/\sigma)}(u_1)\times I_{(-\infty,c+\tau/\sigma)}(u_1)\right)(\sigma u_1-\tau)^2\right]/\sigma^2 \\ &+ 2E\left[(\tilde{\theta} - \theta)'\begin{pmatrix} [I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)](\sigma u_1-\tau) \\ 0 \end{pmatrix}\right]/\sigma^2 \\ &- 2\eta'Z'XS^{-1/2}Q'E\begin{pmatrix} [I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)](\sigma u_1-\tau) \\ 0 \end{pmatrix}/\sigma^2 \end{aligned}$$

$$= \rho(X\tilde{\beta},\ E(y)) + E\left[\left(I_{(-\infty,\tau/\sigma)}(u_1) - I_{(-\infty,c+\tau/\sigma)}(u_1)\right)\right.$$

$$\left.\times(\sigma u_1-\tau)^2\right]/\sigma^2 + 2E\left[\sigma u_1\left(I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)\right)\right.$$

$$\left.\times(\sigma u_1-\tau)\right]/\sigma^2 - 2\xi E\left[\left(I_{(-\infty,c+\tau/\sigma)}(u_1) - I_{(-\infty,\tau/\sigma)}(u_1)\right)\right.$$

$$\left.\times(\sigma u_1-\tau)\right]/\sigma^2$$

$$= \rho(X\tilde{\beta},\ E(y)) - E\left[\left(I_{(-\infty,\tau/\sigma)}(u_1) - I_{(-\infty,c+\tau/\sigma)}(u_1)\right)\right.$$

$$\left.\times(\sigma^2 u_1^2 - \tau^2 - 2\xi(\sigma u_1-\tau))\right]/\sigma^2 \quad . \qquad (4.32)$$

Using simple generalisations of Corollaries 4.1 to 4.3, we obtain the predictive risk function of $\hat{\beta}$ when σ^2 is known as follows:

when $\lambda_1 \le 0$ and $c + \sqrt{2}\lambda_1 \le 0$,

$$\rho(X\hat{\beta},\ E(y)) = k + 2\lambda_2 + \left[\ P(\chi_3^2 \ge (c+\sqrt{2}\lambda_1)^2) - P(\chi_3^2 \ge 2\lambda_1^2)\ \right]/2$$

$$+ \lambda_1^2\left[\ P(\chi_1^2 \ge 2\lambda_1^2) - P(\chi_1^2 \ge (c+\sqrt{2}\lambda_1)^2)\ \right] \quad , \qquad (4.33)$$

when $\lambda_1 > 0$ and $c + \sqrt{2}\lambda_1 \le 0$,

$$\rho(X\hat{\beta},\ E(y)) = k + 2\lambda_2 + \left[\ P(\chi_3^2 \ge (c+\sqrt{2}\lambda_1)^2) + P(\chi_3^2 \ge 2\lambda_1^2)\ \right]/2$$

$$-1 + 2\lambda_1^2 - \lambda_1^2\left[\ P(\chi_1^2 \ge 2\lambda_1^2) + P(\chi_1^2 \ge (c+\sqrt{2}\lambda_1)^2)\ \right] \quad , \qquad (4.34)$$

and when $\lambda_1 > 0$ and $c + \sqrt{2}\lambda_1 > 0$,

$$\rho(X\hat{\beta},\ E(y)) = k + 2\lambda_2 + \left[\ P(\chi_3^2 \ge 2\lambda_1^2) - P(\chi_3^2 \ge (c+\sqrt{2}\lambda_1)^2)\ \right]/2$$

$$+ \lambda_1^2\left[\ P(\chi_3^2 \ge (c+\sqrt{2}\lambda_1)^2) - P(\chi_3^2 \ge 2\lambda_1^2)\ \right] \quad . \qquad (4.35)$$

Note that the case of $\lambda_1 \le 0$ and $c + \sqrt{2}\lambda_1 > 0$ cannot arise as $c < 0$ by assumption. When there is no omitted variable (*i.e.*, the model is correctly specified), $\lambda_2 = 0$ and (4.31) - (4.33) collapse to the expressions given in Judge and Yancey (1986, pp. 97-98) for the weighted risk function of $\hat{\beta}$ with the weight of the squared error loss function equal to X'X.

4.4.2 σ^2 unknown case

The assumption of a known σ^2 is for the purpose of analytical convenience only. In practice, σ^2 is usually unknown and it is therefore necessary to work with the test statistic $t'' = (\tilde{\theta}_1 - r_0)/\tilde{\sigma}$ in testing H_0, where $\tilde{\sigma}^2 = (y - H\tilde{\theta})'(y - H\tilde{\theta})/(n-k)$ is the usual least squares estimator of σ^2. t'' has a doubly non-central t distribution with $v = n - k$ degrees of freedom and non-centrality parameters $\lambda_1^2 = (\tau-\xi)^2/(2\sigma^2)$ and $\lambda_2 = \pi'B'(I-HH')B\pi/(2\sigma^2)$. The applied researcher, unaware of the specification error in the model, believes t'' to have a central t distribution when $\theta = r$ and applies a t test to test the null hypothesis. The researcher's decision rule is to reject the null if $t'' < c$, where c is the size - α critical value for the central t variate with v degrees of freedom and not to reject otherwise.

Again, the case of $c \geq 0$ needs no discussion as the unrestricted estimator is always chosen regardless of the outcome of the test. Following the algebraic manipulation given earlier for the σ^2 known case, when σ^2 is unknown and $c < 0$, the IPTE can be written as :

$$\hat{\beta} = \beta^{**} + S^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{bmatrix}, \tag{4.36}$$

and the corresponding predictive risk function is

$$\rho(X\hat{\beta},\ E(y)) = \rho(X\tilde{\beta},\ E(y)) - E\Big[\Big(I_{(-\infty,\,\tau/\sigma)}(u_1) - I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)\Big) \times(\sigma^2u_1^2 - \tau^2 - 2\xi(\sigma u_1-\tau))\Big]/\sigma^2 \quad . \tag{4.37}$$

The evaluation of this risk is more complicated because it requires the evaluation of $E\left(I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right)$, $j = 0, 1, 2$, which are functions of two stochastic random variables, u_1 and $\tilde{\sigma}$. Now, $E\left(I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right)$ can be written as $E\left(I_{(-\infty,\,cq_v/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right)$, where $q_v = \sqrt{v}\tilde{\sigma}/\sigma$. It is well known (see, for example, Giles and Clarke (1989)) that q_v^2 is distributed as a non-central chi-square variate with v degrees of freedom and non-centrality parameter λ_2; *i.e.*, $q_v^2 \sim \chi'^2_{(v;\lambda_2)}$. In order to evaluate the above expectation, we need Theorem 4.2 which is stated and proved in Appendix 4A. The evaluation of $E\left(I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\right)$ can be undertaken using Corollaries 4.1 to 4.3, as for the case when σ^2 is known.

Theorem 4.2 forms the basis of the following Corollaries :

<u>Corollary 4.4</u>

$$E\left[I_{(-\infty,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)\right] = \begin{cases} \dfrac{E_{1,v}}{2} & \text{if } \lambda_1 \le 0 \\ \\ \dfrac{E_{1,v}}{2} + G_{1,v} & \text{if } \lambda_1 > 0 \end{cases} \quad . \tag{4.38}$$

<u>Corollary 4.5</u>

$$E\left[I_{(-\infty,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1\right] = \begin{cases} -\dfrac{1}{\sqrt{2\pi}}E_{2,v} + \dfrac{\xi}{2\sigma}E_{1,v} & \text{if } \lambda_1 \le 0 \\ \\ -\dfrac{1}{\sqrt{2\pi}}E_{2,v} + \dfrac{\xi}{2\sigma}E_{1,v} + \dfrac{\xi}{\sigma}G_{1,v} & \text{if } \lambda_1 > 0 \end{cases} \quad . \tag{4.39}$$

<u>Corollary 4.6</u>

$$E\left[I_{(-\infty,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^2\right] = \begin{cases} \dfrac{\xi}{2\sigma^2}E_{1,v} - \dfrac{\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}E_{2,v} + E_{3,v}/2 & \text{if } \lambda_1 \le 0 \\ \\ \dfrac{\xi}{2\sigma^2}E_{1,v} + \dfrac{\xi}{\sigma^2}G_{1,v} - \sqrt{\dfrac{2}{\pi}}E_{2,v} & \\ \qquad + E_{3,v}/2 + G_{3,v} & \text{if } \lambda_1 > 0 \end{cases} \tag{4.40}$$

where

$$E_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)} \int_0^{\infty} P(\chi_I^2 \geq (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2(q_J^2)^{J/2+t-1} e^{-q_J^2/2} dq_J^2,$$

$$G_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)} \int_0^{(2v\lambda_1^2/c^2)} P(\chi_I^2 < (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2(q_J^2)^{J/2+t-1} e^{-q_J^2/2} dq_J^2,$$

$I = 1, 2, 3$, and q_J^2 is a Chi-square random variable with J degrees of freedom.

Using Corollaries 4.1 to 4.6, we can show that when $\lambda_1 \leq 0$, the risk of $X\hat{\beta}$ can be written as

$$\rho(X\hat{\beta},\ E(y)) = k + 2\lambda_2 + (E_{3,v} - P_3)/2 - \lambda_1^2(E_{1,v} - P_1) \quad . \tag{4.41}$$

Similarly, when $\lambda_1 > 0$, the risk of $X\hat{\beta}$ is

$$\rho(X\hat{\beta},\ E(y)) = k + 2\lambda_2 - 1 + 2\lambda_1^2 + (E_{3,v} + P_3)/2 - \lambda_1^2(E_{1,v} + P_1)$$
$$- 2\lambda_1^2 G_{1,v} + G_{3,v} \quad . \tag{4.42}$$

When there is no mis-specification in the model, $\lambda_2 = 0$ and (4.41) and (4.42) collapse to expressions equivalent to those given by Hasegawa (1989a).[6]

Again, we perform numerical calculations to examine these risks. They are carried out for n = 10, 30, 50; k = 2, 5; α = 0, 0.01, 0.05, 0.10, 0.25, 0.40; $\lambda_1 \in [-10, 10]$ and various values of λ_2. The NAG (1991) subroutine D01AJF, which is based on an algorithm described in De Doncker (1978), and the subroutines FACTLN and GAMMQ from Press *et al.* (1986), are used to evaluate the integrals $E_{I,J}$ and $G_{I,J}$. For comparison purposes, the risk of $X\hat{\beta}$ for the σ^2 known case is also evaluated. These risk functions are given in Figures 4.1 to 4.6 in Appendix 4C (pp. 106-109).

Although the evaluation of the pre-test predictor's risk is more

---

[6] Hasegawa (1989a) considers the estimation of the mean in a normal population, which is related to, but not identical to, the problem investigated here.

complicated when σ^2 is unknown, the diagrams show that in general, the results are qualitatively the same as when σ^2 is known. For any given λ_2, $\rho(X\hat{\beta}, E(y))$ is bounded and approaches $\rho(X\tilde{\beta}, E(y))$ as $|\lambda_1| \to \infty$. For $\lambda_1 \in (-\infty, 0]$, the inequality $\rho(X\beta^{**}, E(y)) \le \rho(X\hat{\beta}, E(y)) \le \rho(X\tilde{\beta}, E(y))$ always holds. (See Appendix 4A for a proof.[7]) However again, given that λ_1 depends on both τ and ξ, this result alone is not sufficient to ensure the superiority of any estimator over particular intervals of τ. It is possible that the use of the inequality pre-test predictor, even when $\tau \le 0$, can result in a higher risk than would be the case if the unrestricted estimator were used when ξ is sufficiently large.

As $\lambda_2 \to \infty$, $\rho(X\hat{\beta}, E(y))$ is unbounded and approaches[8] $\rho(X\beta^{**}, E(y))$. Figure 4.7 in Appendix 4C illustrates the difference between the risk of the pre-test predictor and that of the unrestricted predictor as λ_2 increases. This diagram shows that over a wide range of the λ_1 space, the risk difference between the two predictors increases as λ_2 increases, typically in the region in which the direction of the constraint is incorrect. Although increasing λ_2 also has the potential for reducing the risk difference, the risk gain is typically very slight compared with the risk loss. It is apparent from these results that if λ_1 and λ_2 are both large, $X\hat{\beta}$ could have infinitely greater risk than the unrestricted predictor $X\tilde{\beta}$. This result is significant as it implies that pre-testing could be potentially dangerous when the errors associated with the

---

[7] This feature is also noted by Judge and Yancey (1986) for the case in which the model is correctly specified. However, they do not provide a formal proof of this result.

[8] This is because for any non-zero c and finite λ_1, $G_{I,v}$ and $E_{I,v}$ both approach zero as $\lambda_2 \to \infty$, i = 1, 2, 3.

constraint and model specification are unknown in practice.[9] This obviously cannot occur if there is no mis-specification in the model, in which case λ_2 vanishes and $\rho(X\hat{\beta}, E(y))$ approaches $\rho(X\tilde{\beta}, E(y))$ as $|\lambda_1| \to \infty$.

Finally, as in the case when the model is properly specified, $\rho(X\hat{\beta}, E(y))$ approaches $\rho(X\tilde{\beta}, E(y))$ and $\rho(X\beta^{**}, E(y))$ as $c \to 0$ or $c \to -\infty$ respectively. The minimum risk boundary of $\rho(X\hat{\beta}, E(y))$ is given by either the risk corresponding to $c \to -\infty$ or the risk corresponding to $c = 0$. Regardless of the size of the pre-test (and hence the value of c), there exists no region in the λ_1 space such that the risk of the pre-test predictor is smaller than the risks of the unrestricted and inequality restricted predictor simultaneously. However, there is always a region such that $X\hat{\beta}$ has higher risk than both $X\tilde{\beta}$ and $X\beta^{**}$. This suggests that if we want to pre-test, then we need to choose an appropriate critical value which brings the pre-test risk function as close as possible to the minimum risk boundary.

4.5 CONCLUSIONS

In this chapter, we have focused on the sampling performance of the inequality restricted and pre-test estimators for the prediction vector in the linear regression model. Our work extends the literature in this area further by allowing for possible model mis-specification through the omission of relevant regressors. It is found that over a relatively large range in the

---

9 In constrast to Figure 4.7, Figure 4.8 (p. 110) shows that over a large portion of the λ_1 space, the percentage risk difference between the pre-test and unrestricted predictors actually decreases as λ_2 increases. This merely indicates that, in this region, the rate of increase of the risk difference between the two predictors, with respect to increasing λ_2, is less than the corresponding rate of increase of the unrestricted predictor's risk. Practically, however, it is the risk difference in absolute terms that is likely to concern the researcher rather than the percentage difference.

(λ_1, λ_2) space, the inequality restricted and pre-test estimators are inferior to the conventional unrestricted estimator in terms of predictive risk.' Even if the inequality constraint is perfectly correct, the use of the inequality restricted estimator may result in a higher risk than would be the case if the unrestricted estimator is used. Both the inequality restricted and pre-test estimators also have the potential to be infinitely worse than the unrestricted estimator in terms of risk when the model is underfitted. These results suggest that while imposing restrictions naively without testing is not recommended, pre-testing is not necessarily the preferred strategy, when the errors associated with the specification of the model and constraint are both unknown in practice. If there is a great deal of uncertainty regarding the correctness of both the inequality restriction and the model specification, the unrestricted estimator may offer a viable and potentially less risky alternative to the inequality restricted and pre-test estimators. The choice of optimal critical values for the pre-test is the subject of the next chapter.

APPENDIX 4A

Theorem 4.1 :

If w is a normal random variable with mean ϑ and variance 1, and $d \in R$, then

$$E\left[I_{(-\infty,d)}(w)w^j\right] = \begin{cases} \sum_{t=0}^{j} \binom{j}{t} (-1)^t \vartheta^{j-t} \Omega_t P(\chi^2_{t+1} \geq f^2)/2 & \text{if } f \leq 0 \quad (4.A.1) \\ \sum_{t=0}^{j} \binom{j}{t} \left[I(t) - P(\chi^2_{t+1} \geq f^2)/2\right] \Omega_t \vartheta^{j-t} & \text{if } f > 0 \quad (4.A.2) \end{cases},$$

where $\Omega_t = 2^{t/2}\Gamma((t+1)/2)/\Gamma(\frac{1}{2})$, $f = d-\vartheta$ and $I(t) = 0$ if t is odd, 1 otherwise. This theorem generalises Theorem 1 of Judge and Yancey (1986, pp.72-73) to a normal variable with a non-zero mean.

Proof :

Let $z = (w - \vartheta) \sim N(0,1)$. Then $E\left[I_{(-\infty,d)}(w)w^j\right]$ can be written as $E\left[I_{(-\infty,d-\vartheta)}(z)(z+\vartheta)^j\right]$. Now $E\left[I_{(-\infty,d-\vartheta)}(z)(z^t\vartheta^{j-t})\right]$, $t = 0, 1,..,j$, can be evaluated using Theorem 1 of Judge and Yancey (1986). Theorem 4.1 then follows.

Theorem 4.2 :

Let w, d, f and I(t) be defined as in Lemma 1. Let $\psi \sim \chi'^2_{(v;\lambda)}$ and $c \in R^-$, then for $f \leq 0$,

$$E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]$$
$$= \sum_{t=0}^{j} \binom{j}{t} (-1)^t \left\{\vartheta^{j-t}\Omega_t e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!\,2^{v/2+i}\Gamma(\frac{v}{2}+i)} \left[\int_0^\infty P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right) \psi^{\frac{\nu}{2}+i-1} e^{-\frac{\psi}{2}}/2 \, d\psi \right.\right.$$

(4.A.3)

and for f > 0,

$$E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]$$

$$= \sum_{t=0}^{j}\binom{j}{t}(-1)^t\left\{\vartheta^{j-t}\Omega_t e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!2^{v/2+i}\Gamma(\frac{v}{2}+i)}\left[\int_0^{\infty}P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)\psi^{\frac{\nu}{2}+i-1}e^{-\frac{\psi}{2}}/2\ d\psi\right.\right.$$

$$\left.\left.+I(t)\int_0^{f^2/c^2}P\left(\chi^2_{t+1}<(c\sqrt{\psi}+f)^2\right)\psi^{\frac{\nu}{2}+i-1}e^{-\frac{\psi}{2}}d\psi\right]\right\} \quad . \qquad (4.A.4)$$

Proof :

As $c \in R^-$ and $\sqrt{\psi} \geq 0$ by definition, we have $f \leq 0$ implying that $c\sqrt{\psi}+f \leq 0$. Using Theorem 4.1, $E_{w|c\sqrt{\psi}+d\leq 0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t$ $P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)/2$. Therefore, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t$ $E_{\sqrt{\psi}}\left[P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)\right]/2$ when $f \leq 0$. Now, $P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)$ is a function of $\sqrt{\psi}$, and $\sqrt{\psi}$ is defined only on the non-negative horizon such that each ψ corresponds to a unique $\sqrt{\psi}$. Hence $P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)$ can be regarded as a function of ψ. Therefore, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t\int_0^{\infty}P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)/2\ \ p(\psi)\ d\psi$ when $f \leq 0$, which leads to (4.A.3). Now, when $f > 0$, the sign of $c\sqrt{\psi} + f$ is undetermined. When $c\sqrt{\psi} + f \leq 0$, the range of ψ is restricted to $\psi \geq f^2/c^2$, while $c\sqrt{\psi} + f > 0 \Rightarrow 0 < \psi < f^2/c^2$. Hence when $f > 0$, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = E\left\{I_{(f^2/c^2,\infty)}(\psi)E_{w|c\sqrt{\psi}+f\leq 0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] + I_{(0,f^2/c^2)}(\psi)E_{w|c\sqrt{\psi}+f>0}\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right]\right\}$. The two inner expectations may be evaluated using Theorem 4.1, giving $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = E\left\{I_{(f^2/c^2,\infty)}(\psi)\sum_{t=0}^{j}\right.$ $\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t P\left(\chi^2_{t+1}\geq(c\sqrt{\psi}+f)^2\right)/2 + I_{(0,f^2/c^2)}(\psi)\sum_{t=0}^{j}\binom{j}{t}\Big[I(t) -$

$P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right)/2\Big]\vartheta^{j-t}\Omega_t\Big\}$. Now, using the fact that $I_{(f^2/c^2,\infty)}(\psi) = 1 - I_{(0,f^2/c^2)}(\psi)$, and after some manipulations, we can show that $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t E_\psi\Big\{\left[P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right)\right]/2 + I(t)\, I_{(0,f^2/c^2)}(\psi) - \left((-1)^t+1\right) I_{(0,f^2/c^2)}(\psi)\ P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right)/2\Big\}$. As $((-1)^t + 1) = 0$ if t is odd, 2 otherwise, then $((-1)^t + 1) = 2I(t)$. Therefore, $E\left[I_{(-\infty,c\sqrt{\psi}+d)}(w)w^j\right] = \sum_{t=0}^{j}\binom{j}{t}(-1)^t\vartheta^{j-t}\Omega_t E_\psi\Big\{\left[P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right)\right]/2 + I(t)\, I_{(0,f^2/c^2)}(\psi)\left[1-P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right)\right]\Big\}$. Finally, noting that $E_\psi\left[I_{(0,f^2/c^2)}(\psi)F(\psi)\right] = \int_0^{f^2/c^2} F(\psi)\ p(\psi)\ d\psi$ and $1 - P\left(\chi^2_{t+1} \geq (c\sqrt{\psi}+f)^2\right) = P\left(\chi^2_{t+1} < (c\sqrt{\psi}+f)^2\right)$ for any given ψ, (4.A.4) follows directly.

Lemma 4.1 :

$\rho(X\beta^{**}, E(y)) \leq \rho(X\hat{\beta}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$ when $\lambda_1 \leq 0$

Proof :

In comparing (4.24) and (4.39), $\rho(X\beta^{**}, E(y)) \leq \rho(X\hat{\beta}, E(y))$ *if and only if*

$$E_{3,v} \geq 2\lambda_1^2 E_{1,v}\ . \tag{4.A.5}$$

Now, for any given q_v, (4.A.5) is true *if and only if*

$$P\left(\chi_3^2 \geq d^2\right) \geq 2\lambda_1^2 P\left(\chi_1^2 \geq d^2\right) \tag{4.A.6}$$

where $d^2 = (cq_v/\sqrt{v} + \sqrt{2}\lambda_1)^2$.

Now, we know that for any given q_v,

$$P\left(\chi_1^2 \geq d^2\right) = \int_{d^2}^{\infty} \frac{\psi^{-1/2}\exp(-\psi/2)}{\sqrt{2}\ \Gamma(\frac{1}{2})}\, d\psi \tag{4.A.7a}$$

and

$$P\left(\chi_3^2 \geq d^2\right) = \int_{d^2}^{\infty} \frac{\psi^{1/2}\exp(-\psi/2)}{2^{3/2}\,\Gamma(\frac{3}{2})}\,d\psi, \qquad (4.A.7b)$$

where $\psi = z^2$ and z is a standard normal random variable.

In order to prove (4.A.6), we need the following theorem :

Theorem 4.3 :

If f(x) is positive and continuous and $0 \leq c \leq c_1 \leq c_2$, then

$$c\int_{c_1}^{c_2} f(x)\,dx \leq \int_{c_1}^{c_2} xf(x)\,dx \qquad (4.A.8)$$

Proof :

As $c \leq c_1 \leq c_2$ and f(x) is positive and continuous, it is obvious that $xf(x) \geq cf(x)$ for all $x \in [c_1, c_2]$. Accordingly, (4.A.8) must hold.

Q.E.D.

Now, when $\lambda_1 \leq 0$, $2\lambda_1^2$ must not be greater than d^2 as $c < 0$ and q_v only takes on positive values. Using this fact and Theorem 4.1, we establish the following :

$$2\lambda_1^2\int_{d^2}^{\infty} \frac{\psi^{-1/2}\exp(-\psi/2)}{\sqrt{2}\,\Gamma(\frac{1}{2})}\,d\psi \leq \int_{d^2}^{\infty} \frac{\psi\,\psi^{-1/2}\exp(-\psi/2)}{\sqrt{2}\,\Gamma(\frac{1}{2})}\,d\psi = \int_{d^2}^{\infty} \frac{\psi^{1/2}\exp(-\psi/2)}{2^{3/2}\Gamma(\frac{3}{2})}\,d\psi$$

$$= P\left(\chi_3^2 \geq d^2\right).$$

Therefore, for any given q_v,

$$P\left(\chi_3^2 \geq d^2\right) \geq 2\lambda_1^2 P\left(\chi_1^2 \geq d^2\right) \qquad (4.A.9)$$

which implies

$$E_{3,v} \geq 2\lambda_1^2 E_{1,v}\ . \qquad (4.A.10)$$

Hence $\rho(X\beta^{**}, E(y)) \le \rho(X\hat{\beta}, E(y))$.

To show $\rho(X\hat{\beta}, E(y)) \le \rho(X\tilde{\beta}, E(y))$ when $\lambda_1 \le 0$, we require

$$E_{3,v} - 2\lambda_1^2 E_{1,v} \le P_3 - 2\lambda_1^2 P_1 \ . \tag{4.A.11}$$

Now,

$$E_{3,v} - 2\lambda_1^2 E_{1,v}$$

$$= e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{\infty} \int_{d^2}^{\infty} (\psi - 2\lambda_1^2) \frac{1}{\sqrt{2\pi}} \frac{\exp(-\psi/2)}{\sqrt{\psi}} \, d\psi \, (q_v^2)^{v/2+t-1} e^{-q_v^2/2} \, dq_v^2, \tag{4.A.12}$$

as $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Also,

$$P_3 - 2\lambda_1^2 P_1$$

$$= \int_{2\lambda_1^2}^{\infty} (\psi - 2\lambda_1^2) \frac{1}{\sqrt{2\pi}} \frac{\exp(-\psi/2)}{\sqrt{\psi}} \, d\psi$$

$$= \int_{2\lambda^2}^{\infty} (\psi - 2\lambda_1^2) \frac{1}{\sqrt{2\pi}} \frac{\exp(-\psi/2)}{\sqrt{\psi}} \, d\psi \times e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{\infty} (q_v^2)^{v/2+t-1} e^{-q_v^2/2} \, dq_v^2,$$

as $e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{\infty} (q_v^2)^{v/2+t-1} e^{-q_v^2/2} \, dq_v^2 = 1.$

Therefore,

$$P_3 - 2\lambda_1^2 P_1$$

$$= e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{v/2+t}\Gamma(\frac{v}{2}+t)} \int_0^{\infty} \int_{2\lambda}^{\infty} (\psi - 2\lambda_1^2) \frac{1}{\sqrt{2\pi}} \frac{\exp(-\psi/2)}{\sqrt{\psi}} \, d\psi \, (q_v^2)^{v/2+t-1} e^{-q_v^2/2} \, dq_v^2 \ . \tag{4.A.13}$$

Now, for any particular value of q_v, $(\psi - 2\lambda_1^2)\ \dfrac{1}{\sqrt{2\pi}}\dfrac{\exp(-\psi/2)}{\sqrt{\psi}}\ (q_v^2)^{v/2+t-1}\ e^{-q_v^2/2} \geq 0$ for $\psi \in [2\lambda_1^2, \infty)$ and $d^2 \geq 2\lambda_1^2$ when $\lambda_1 \leq 0$. From these results, it is clear that $P_3 - 2\lambda_1^2 P_1 \geq E_{3,v} - 2\lambda_1^2 E_{1,v}$, which implies $\rho(X\hat{\beta}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$.

We have shown $\rho(X\beta^{**}, E(y)) \leq \rho(X\hat{\beta}, E(y))$ and $\rho(X\hat{\beta}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$, hence the inequality $\rho(X\beta^{**}, E(y)) \leq \rho(X\hat{\beta}, E(y)) \leq \rho(X\tilde{\beta}, E(y))$ when $\lambda_1 \leq 0$ is established.

APPENDIX 4B : THE RISK FUNCTIONS OF β^{**} and $\hat{\beta}$

For purposes of completeness and general use, we also derive the risk of β^{**} and $\hat{\beta}$ in addition to the risks of $X\beta^{**}$ and $X\hat{\beta}$ given earlier.

Under a squared error loss measure, the risk of β^{**}, relative to σ^2, may be expressed as

$$\begin{aligned}
\rho(\beta^{**}, \beta) &= E\left[(\beta^{**} - \beta)'(\beta^{**} - \beta)\right]/\sigma^2 \\
&= E\left[\left(\tilde{\beta} - \beta - S^{-1/2}Q'\begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right)' \right. \\
&\quad \left. \cdot \left(\tilde{\beta} - \beta - S^{-1/2}Q'\begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right)\right]/\sigma^2 \\
&= \rho(\tilde{\beta}, \beta) \\
&\quad + E\left\{\begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}' QS^{-1}Q' \begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right\}/\sigma^2 \\
&\quad - 2E\left\{(\tilde{\beta} - \beta)'S^{-1/2}Q'\begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right\}/\sigma^2 \quad . \qquad (4.B.1)
\end{aligned}$$

Now, if we let $QS^{-1}Q' = A = \begin{pmatrix} a_1 & a_3' \\ a_3 & a_2 \end{pmatrix}$, where a_1 is 1 x 1, a_2 is $(k - 1) \times (k - 1)$ and a_3 is $(k - 1) \times 1$, then the risk of β^{**} can be written as

$$\begin{aligned}
\rho(\beta^{**}, \beta) &= \rho(\tilde{\beta}, \beta) + a_1 E\left[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau)^2\right]/\sigma^2 \\
&\quad - 2E\left[(\tilde{\theta} - \theta)'\begin{pmatrix} a_1 & a_3' \\ a_3 & a_2 \end{pmatrix}\begin{pmatrix} I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right]/\sigma^2 \\
&= \rho(\tilde{\beta}, \beta) + a_1 E\left[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau)^2\right]/\sigma^2 \\
&\quad - 2a_1 E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^2\right] - 2a_1\tau E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1\right]/\sigma \\
&\quad - 2E\left[u'_{(k-1)}a_3 I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau)\right]/\sigma \\
&= \rho(\tilde{\beta}, \beta) - a_1 E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^2\right] + a_1 E\left[I_{(-\infty,\tau/\sigma)}(u_1)\tau^2\right]/\sigma^2 \\
&\quad - 2E\left[u'_{(k-1)}a_3 I_{(-\infty,\tau/\sigma)}(u_1)(\sigma u_1 - \tau)\right]/\sigma \qquad (4.B.2)
\end{aligned}$$

where $u'_{(k-1)}$ is the vector of the last k-1 elements in $(\tilde{\theta}-\theta)/\sigma$.

Recognising that $\rho(\tilde{\beta}, \beta) = trS^{-1} + \eta'Z'XS^{-2}X'Z\eta/\sigma^2$ and using Corollaries

4.1 to 4.3, the inequality restricted estimator's risk function may be expressed, when $\lambda_1 \le 0$, as

$$\rho(\beta^{**}, \beta) = \mathrm{tr}S^{-1} + \eta' Z' X S^{-2} X' Z\eta/\sigma^2 - a_1(\xi^2-\tau^2)P(\chi_1^2 \ge 2\lambda_1^2)/(2\sigma^2)$$
$$+ a_1 \frac{\xi}{\sigma}\sqrt{\frac{2}{\pi}}P(\chi_2^2 \ge 2\lambda_1^2) - a_1 P(\chi_3^2 \ge 2\lambda_1^2)/2$$
$$+ 2(QS^{-1/2}X'Z\eta)'_{(k-1)} a_3\left[\lambda_1 P(\chi_1^2 \ge 2\lambda_1^2)/2 + P(\chi_2^2 \ge 2\lambda_1^2)/\sqrt{2\pi}\right], \tag{4.B.3}$$

or, when $\lambda_1 > 0$, as

$$\rho(\beta^{**}, \beta) = \mathrm{tr}S^{-1} + \eta' Z' X S^{-2} X' Z\eta/\sigma^2 - a_1(\xi^2-\tau^2)/\sigma^2$$
$$+ a_1(\xi^2-\tau^2)P(\chi_1^2 \ge 2\lambda_1^2)/(2\sigma^2) + a_1 \frac{\xi}{\sigma}\sqrt{\frac{2}{\pi}}P(\chi_2^2 \ge 2\lambda_1^2) - a_1$$
$$+ a_1 P(\chi_3^2 \ge 2\lambda_1^2)/2 + 2(QS^{-1/2}X'Z\eta)'_{(k-1)} a_3\Big[\lambda_1$$
$$- \lambda_1 P(\chi_1^2 \ge 2\lambda_1^2)/2 + P(\chi_2^2 \ge 2\lambda_1^2)/\sqrt{2\pi}\Big]. \tag{4.B.4}$$

Similarly, using a squared error loss measure, the risk of $\hat{\beta}$, relative to σ^2, may be written as:

$$\begin{aligned}
\rho(\hat{\beta}, \beta) &= E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]/\sigma^2 \\
&= E\left[\left(\beta^{**} - \beta + S^{-1/2}Q'\begin{pmatrix} I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right)'\right. \\
&\quad \times \left.\left(\beta^{**} - \beta + S^{-1/2}Q'\begin{pmatrix} I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right)'\right]/\sigma^2 \\
&= \rho(\beta^{**}, \beta) + E\left\{\begin{pmatrix} I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}'\right. \\
&\quad \times \left. QS^{-1}Q'\begin{pmatrix} I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right\}/\sigma^2 \\
&\quad + 2E\left\{(\beta^{**} - \beta)'S^{-1/2}Q'\begin{pmatrix} I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right\}/\sigma^2 \\
&= \rho(\beta^{**}, \beta) + a_1 E\left[I_{(-\infty, (c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1 - \tau)^2\right]/\sigma^2 \\
&\quad + 2E\left\{\left[\tilde{\beta} - \beta - S^{-1/2}Q'\begin{pmatrix} I_{(-\infty, \tau/\sigma)}(u_1)(\sigma u_1 - \tau) \\ 0 \end{pmatrix}\right]'\right.
\end{aligned}$$

$$\times\ S^{-1/2}Q'\begin{pmatrix} I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix}\Bigg\}\Big/\sigma^2$$

$$= \rho(\beta^{**},\ \beta) + a_1 E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau)^2\right]\Big/\sigma^2$$

$$+\ 2E\left[(\tilde{\theta} - \theta)'\begin{pmatrix} a_1 & a_3' \\ a_3 & a_2 \end{pmatrix}\begin{pmatrix} I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix}\right]\Big/\sigma^2$$

$$-\ 2E\left\{\begin{pmatrix} I_{(-\infty,\,\tau/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix}'\begin{pmatrix} a_1 & a_3' \\ a_3 & a_2 \end{pmatrix}\right.$$

$$\left.\times\begin{pmatrix} I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau) \\ 0 \end{pmatrix}\right\}\Big/\sigma^2$$

$$= \rho(\beta^{**},\ \beta) + a_1 E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau)^2\right]\Big/\sigma^2$$

$$+\ 2a_1 E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^2\right]$$

$$-\ 2a_1\tau E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1\right]\Big/\sigma$$

$$+\ 2E\left[u'_{(k-1)}a_3 I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau)\right]\Big/\sigma$$

$$-\ 2a_1 E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau)^2\right]\Big/\sigma^2$$

$$= \rho(\beta^{**},\ \beta) + a_1 E\left[I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma^2u_1^2-\tau^2)\right]$$

$$+\ 2E\left[u'_{(k-1)}a_3 I_{(-\infty,\,(c\tilde{\sigma}+\tau)/\sigma)}(u_1)(\sigma u_1-\tau)\right]\Big/\sigma \qquad . \tag{4.B.4}$$

Using Corollaries 4.1 to 4.6, we may express the risk of $\hat{\beta}$, when $\lambda_1 \le 0$, as

$$\rho(\hat{\beta},\ \beta) = \rho(\beta^{**},\ \beta) + a_1(\xi^2-\tau^2)E_{1,v}/(2\sigma^2) - a_1\xi\sqrt{\frac{2}{\pi}}\ E_{2,v}/\sigma + a_1E_{3,v}/2$$

$$-\ 2(QS^{-1/2}X'Z\eta)'_{(k-1)}a_3\left[E_{2,v}/\sqrt{2\pi} + \lambda_1E_{1,v}/2\right]. \tag{4.B.5}$$

Alternatively, when $\lambda_1 > 0$, the risk of $\hat{\beta}$ may be expressed as

$$\rho(\hat{\beta},\ \beta) = \rho(\beta^{**},\ \beta) + a_1(\xi^2-\tau^2)E_{1,v}/(2\sigma^2) - a_1\xi\sqrt{\frac{2}{\pi}}\ E_{2,v}/\sigma + a_1E_{3,v}/2$$

$$-\ 2\left\{(QS^{-1/2}X'Z\eta)'_{(k-1)}a_3E_{2,v}/\sqrt{2\pi} + \lambda_1E_{1,v}/2\right\}.$$

$$-\ a_1(\xi^2-\tau^2)G_{1,v}/\sigma^2 - a_1G_{3,v} + 2\left\{(QS^{-1/2}X'Z\eta)'_{(k-1)}a_3\lambda_1G_{1,v}\right\}.$$

$$\tag{4.B.6}$$

Clearly, the risks of β^{**} and $\hat{\beta}$ are data dependent. For the purposes of illustrating our analysis, several data series have been used to evaluate these expressions. These data series are, the annual price and income series from Durbin and Watson's (1951) consumption of spirits example; and the monthly unemployment rate in New Zealand (December 1985 - September 1990).

These risks are shown in Figures 4.9 to 4.12 (pp. 111-112) as functions of τ, the surplus variable. The included regressors consist of an intercept, and either the unemployment rate series or the income series from Durbin and Watson's (1951) consumption of spirits example. The regressors omitted from the model are either seasonal dummy variables, or the price series from Durbin and Watson's (1951) data.[10] At least for the cases that we have considered, the risk of the estimators for the coefficient vector depict essentially the same characteristics as the risks of the prediction vector. When the model is correctly specified, both the IRE and IPTE are risk superior to the unrestricted estimator when $\tau \leq 0$. The risk of the IPTE is always no less than the minimum of the risk of the UE and IRE.

When the model is mis-specified, depending on the data and the magnitude of the omitted regressors' coefficients, the risk functions either shift to the left or to the right of their correctly specified counterparts. Once again, this illustrates that in an underfitted model, the use of valid prior information does not necessarily lead to the reduction in risk. At least for the cases that we have considered, the IPTE is never superior to both the IRE and the unrestricted estimator, as in the case when the model is properly specified. Over a wide range in the parameter space, the absolute risk difference between the risk of the unrestricted and the inequality pre-test

---

[10] The omission of seasonal dummies does not apply to Durbin and Watson's (1951) data as it is annual data.

estimators increases with the degree of model mis-specification. As we have considered only a small range of data series, and given that these risk functions are data dependent, more investigations are necessary before the impact of excluding relevant regressors on the risk of the IRE and IPTE of the coefficient vector is fully understood.

APPENDIX 4C

Figure 4.1

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for $n = 30$, $k = 5$ and $\lambda_2 = 0$ (σ^2 known case)

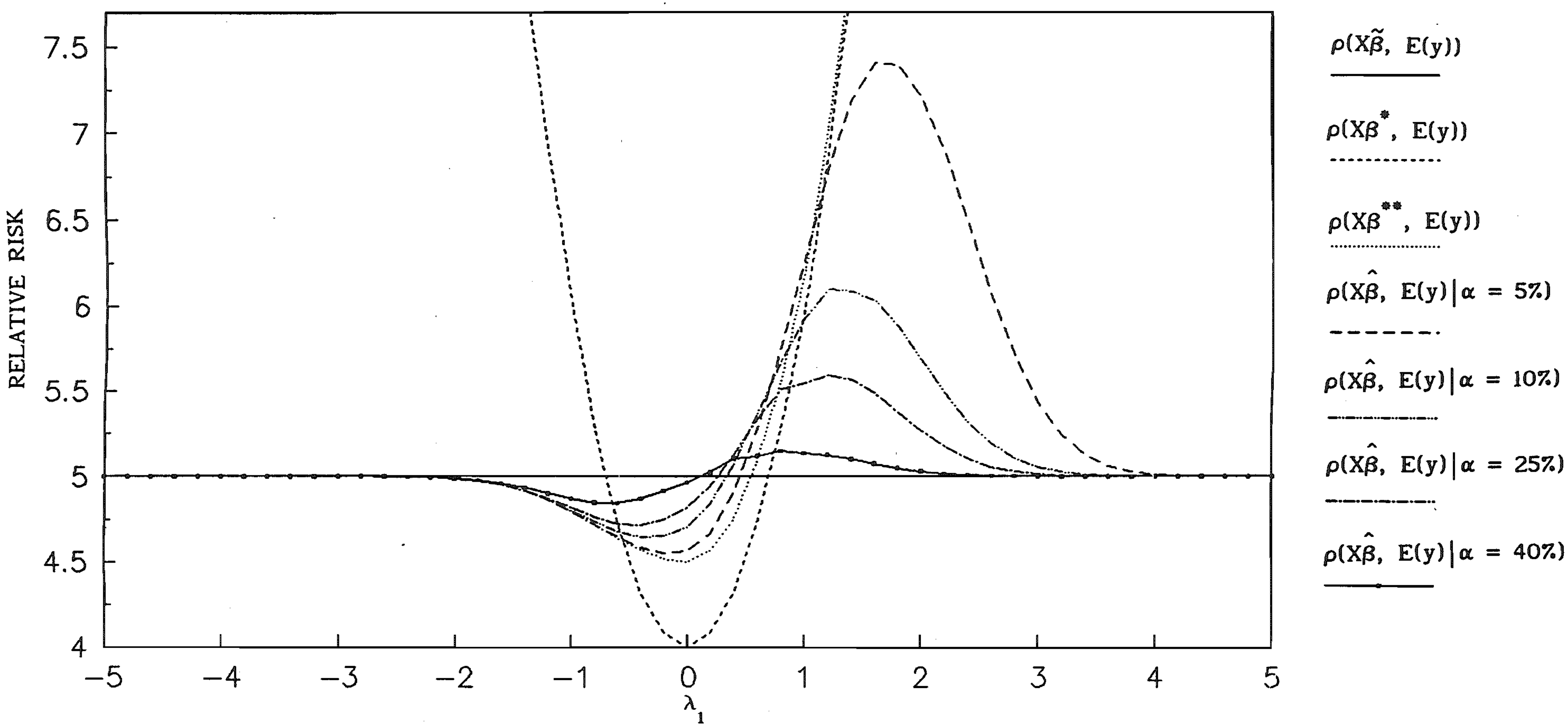


Figure 4.2

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for n = 30, k = 5 and λ_2 = 2 (σ^2 known case)

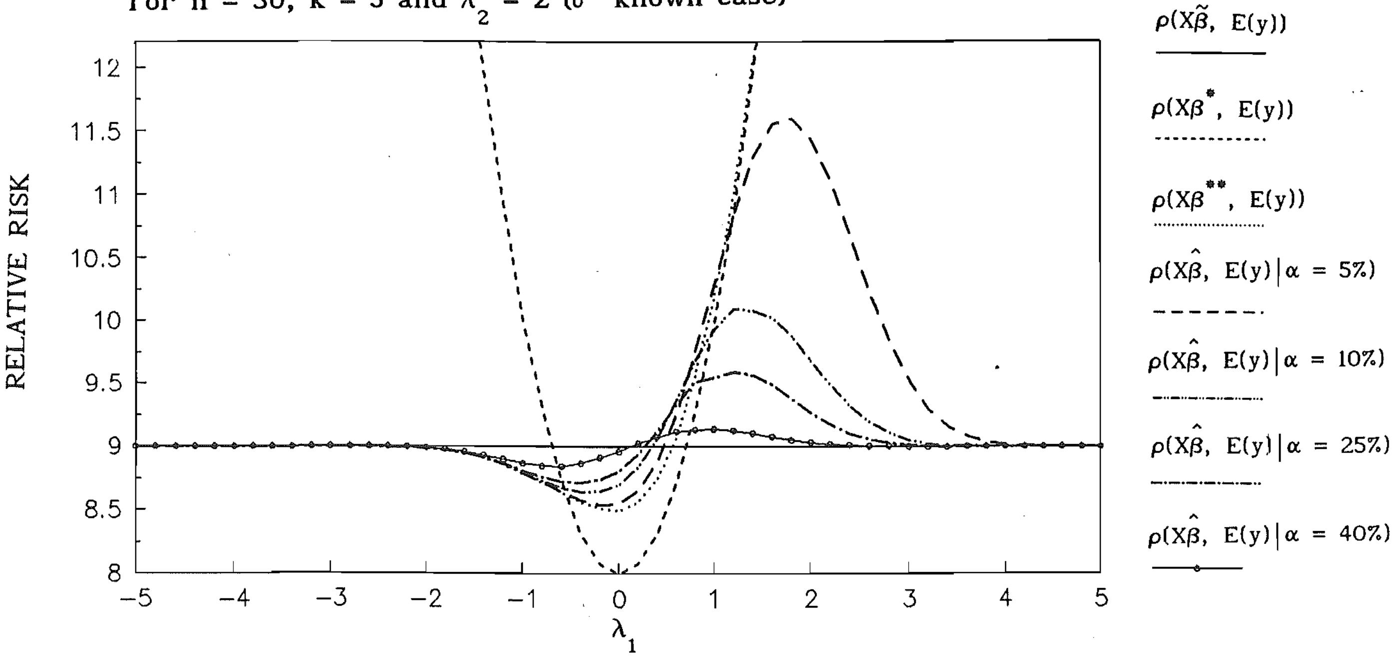


Figure 4.3

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for n = 30, k = 5 and λ_2 = 10 (σ^2 known case)

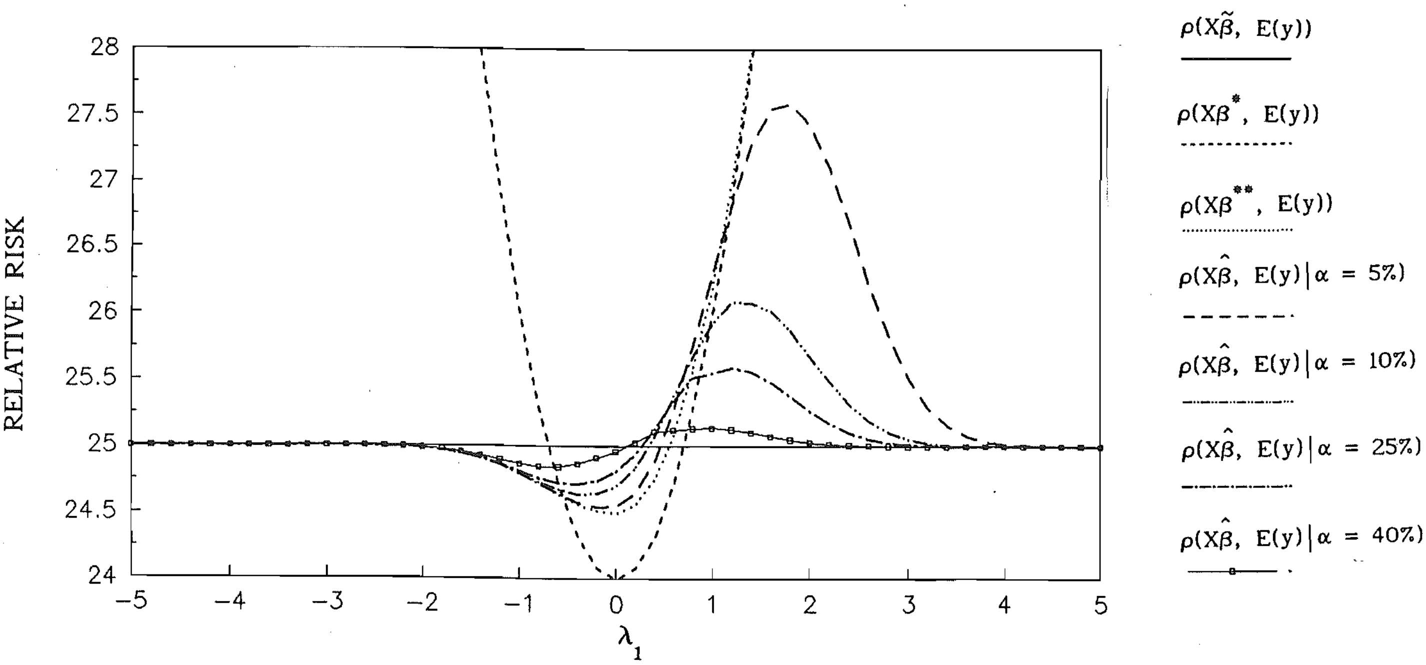


Figure 4.4

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for $n = 30$, $k = 5$ and $\lambda_2 = 0$ (σ^2 unknown case)

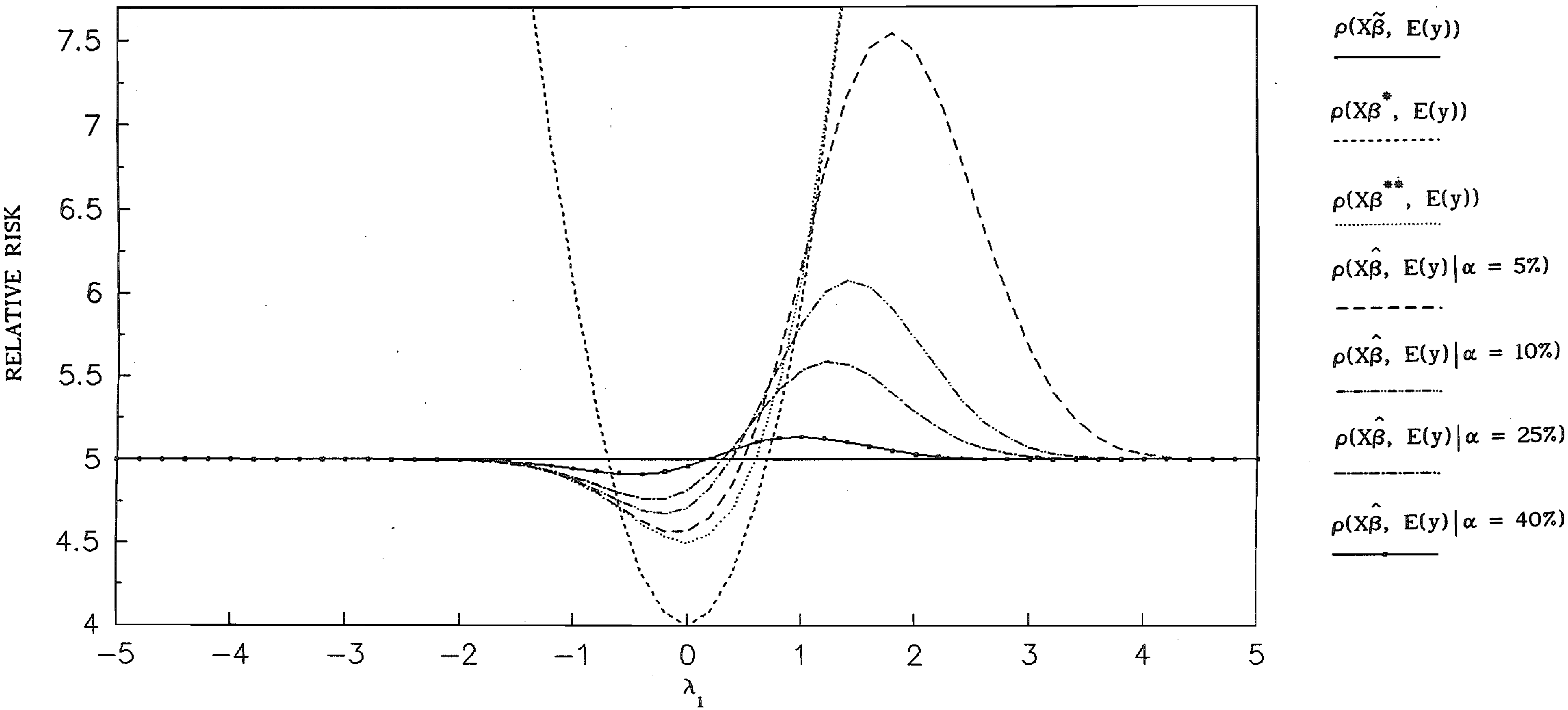


Figure 4.5

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for n = 30, k = 5 and λ_2 = 2 (σ^2 unknown case)

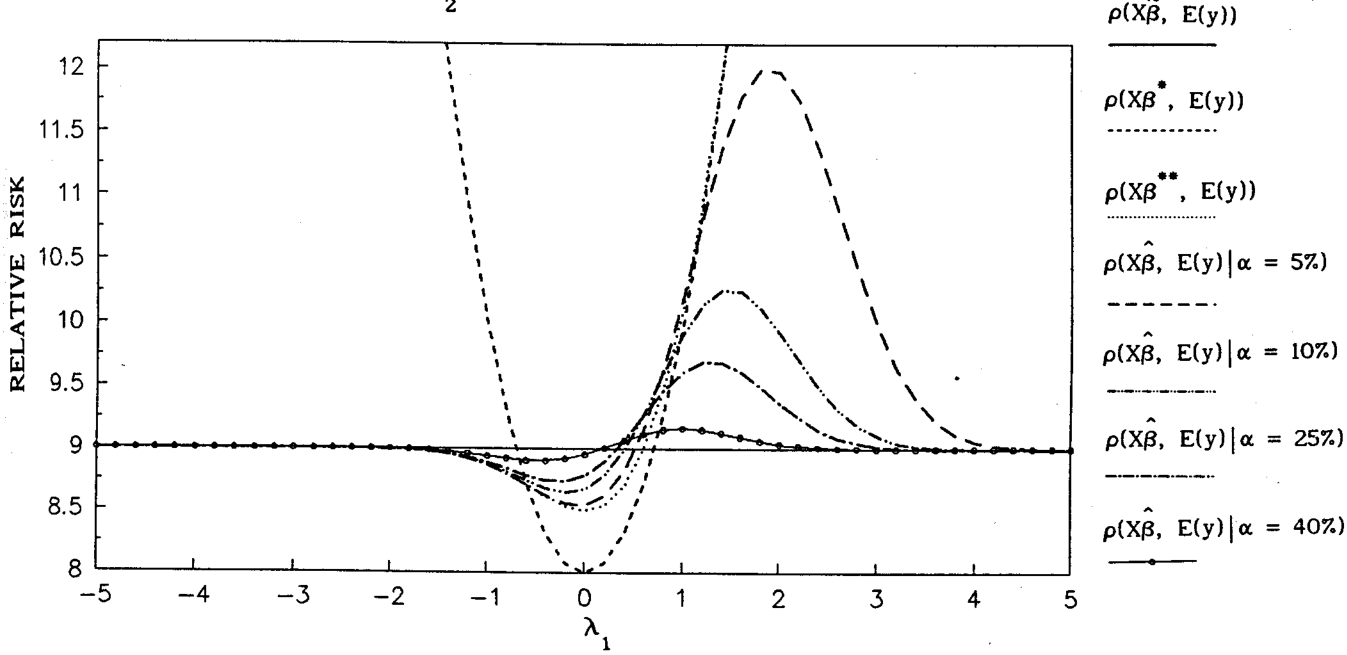


Figure 4.6

Relative risk functions of $X\tilde{\beta}$, $X\beta^{*}$, $X\beta^{**}$ and $X\hat{\beta}$ for n = 30, k = 5 and λ_2 = 10 (σ^2 unknown case)

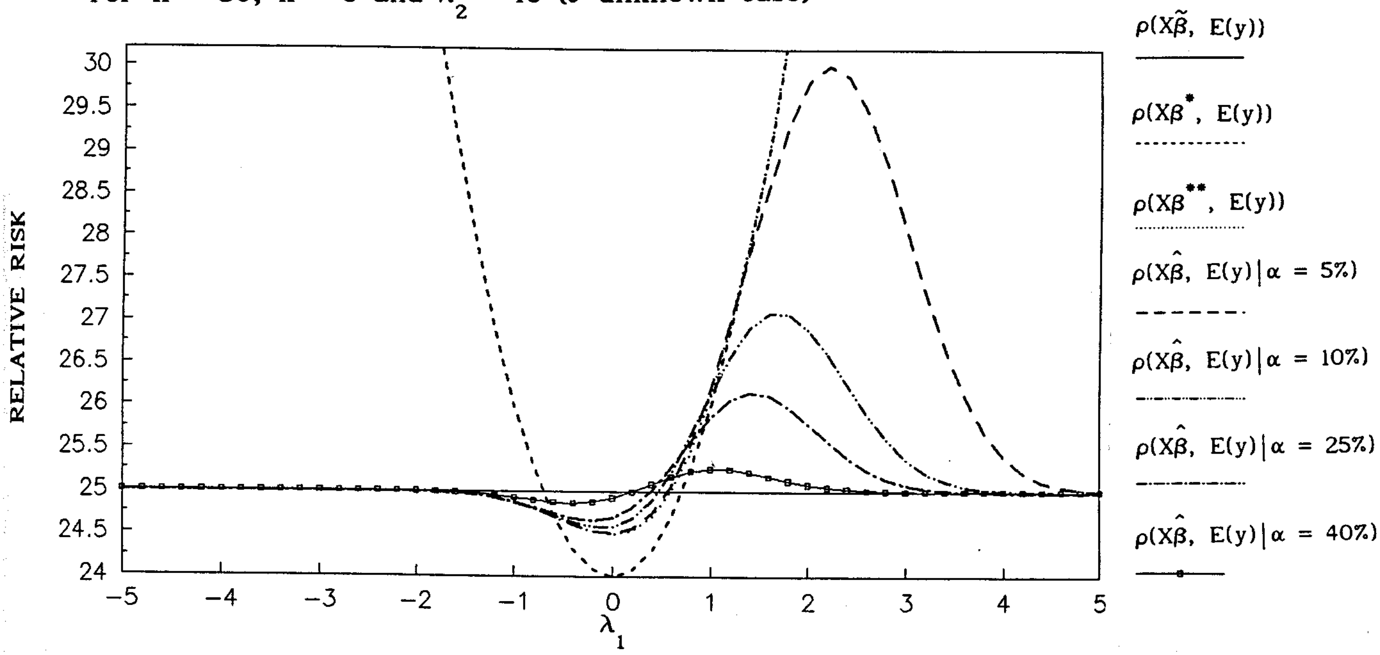


Figure 4.7

Risk difference between $X\hat{\beta}$ and $X\tilde{\beta}$ for n = 30, k = 5 and α = 0.01

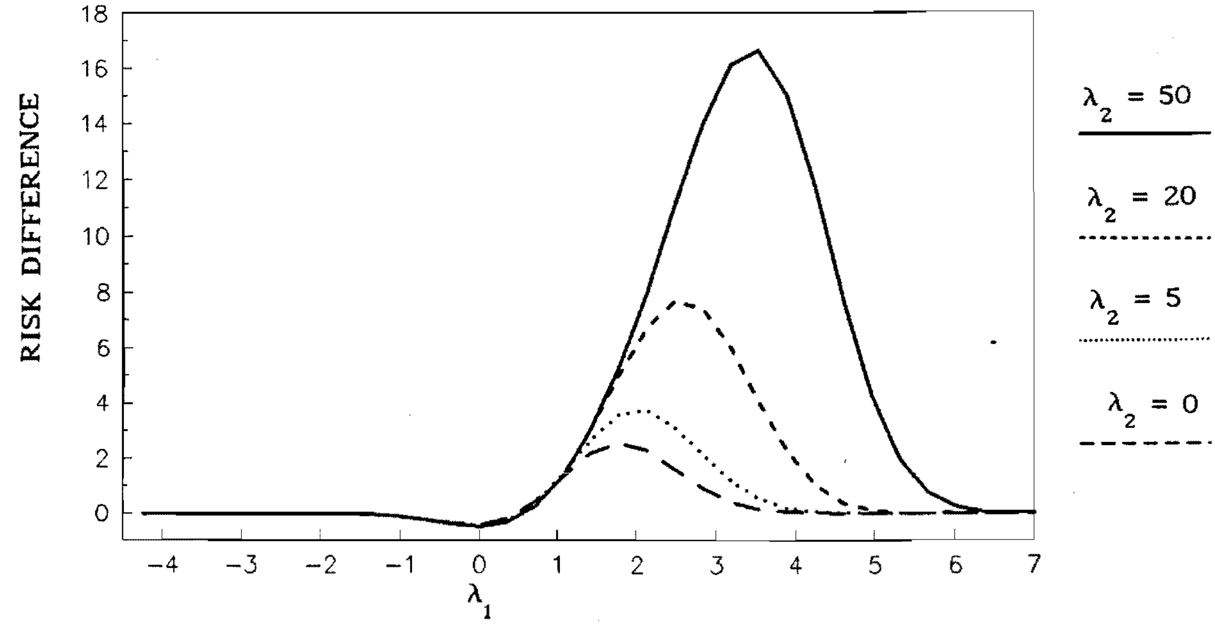


Figure 4.8

Percentage difference between the risk of $X\tilde{\beta}$ and $X\hat{\beta}$ for n = 30, k = 5 and α = 0.01

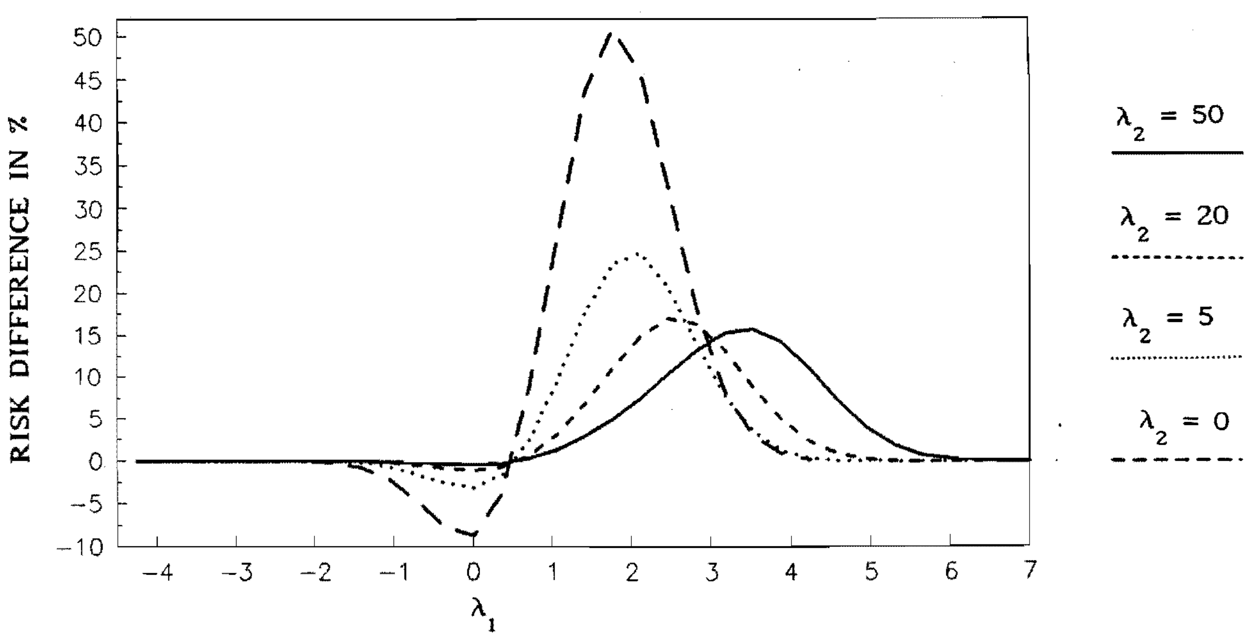


Figure 4.9

Relative risk functions of $\tilde{\beta}$, β^*, β^{**} and $\hat{\beta}$ in a properly specified model
unemployment data : n = 20

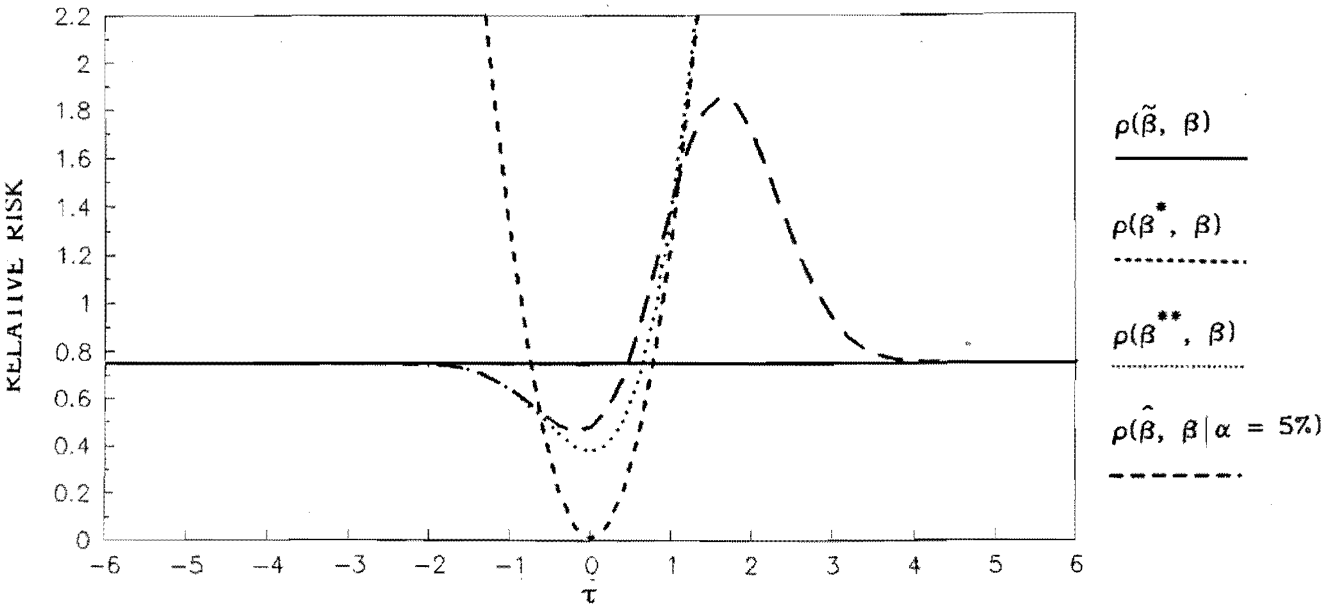


Figure 4.10

Relative risk functions of $\tilde{\beta}$, β^*, β^{**} and $\hat{\beta}$ in a mis-specified model

unemployment data : n = 20
omitted regressors : seasonal dummies (β_3 = -2.5, β_4 = 1.6 and β_5 = 3.8)

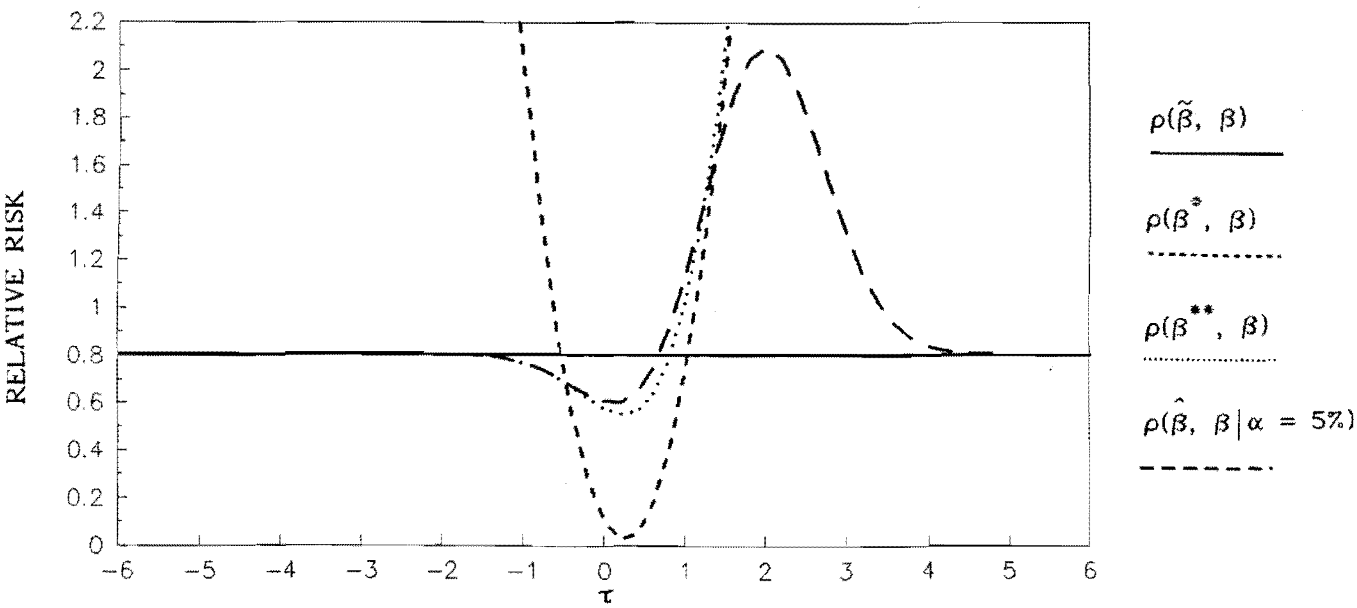


Figure 4.11

Relative risk functions of $\tilde{\beta}$, β^{*}, β^{**} and $\hat{\beta}$ in a properly specified model income series from spirits data: n = 60

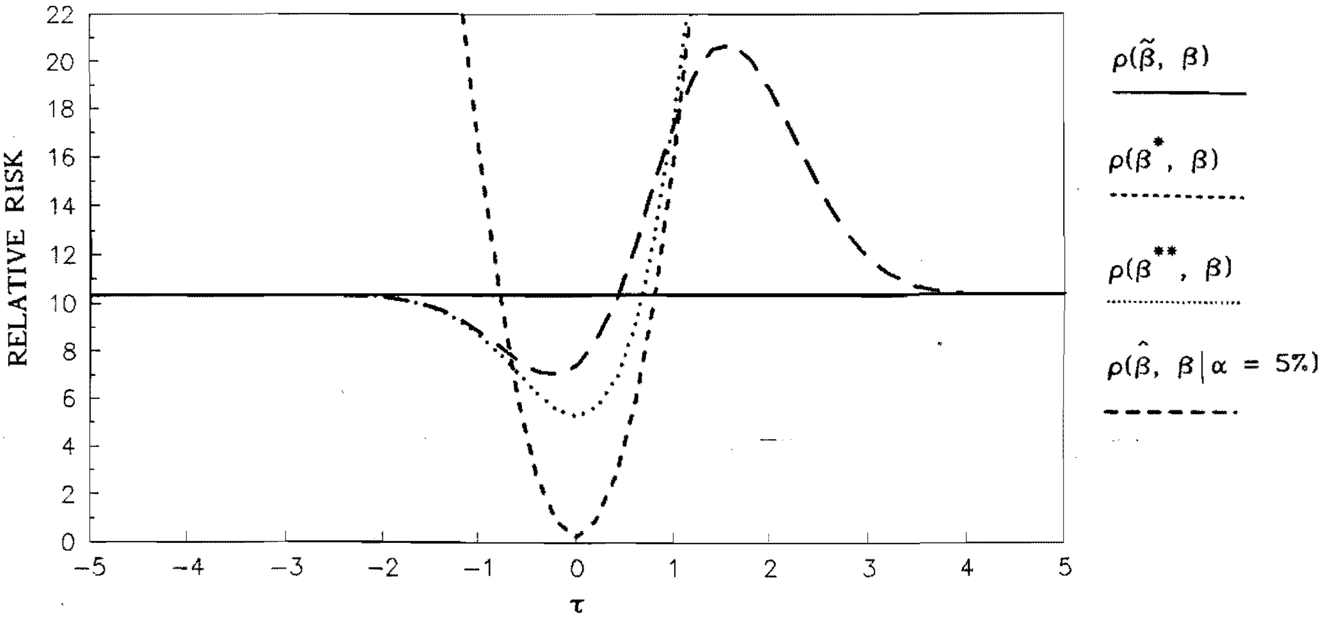
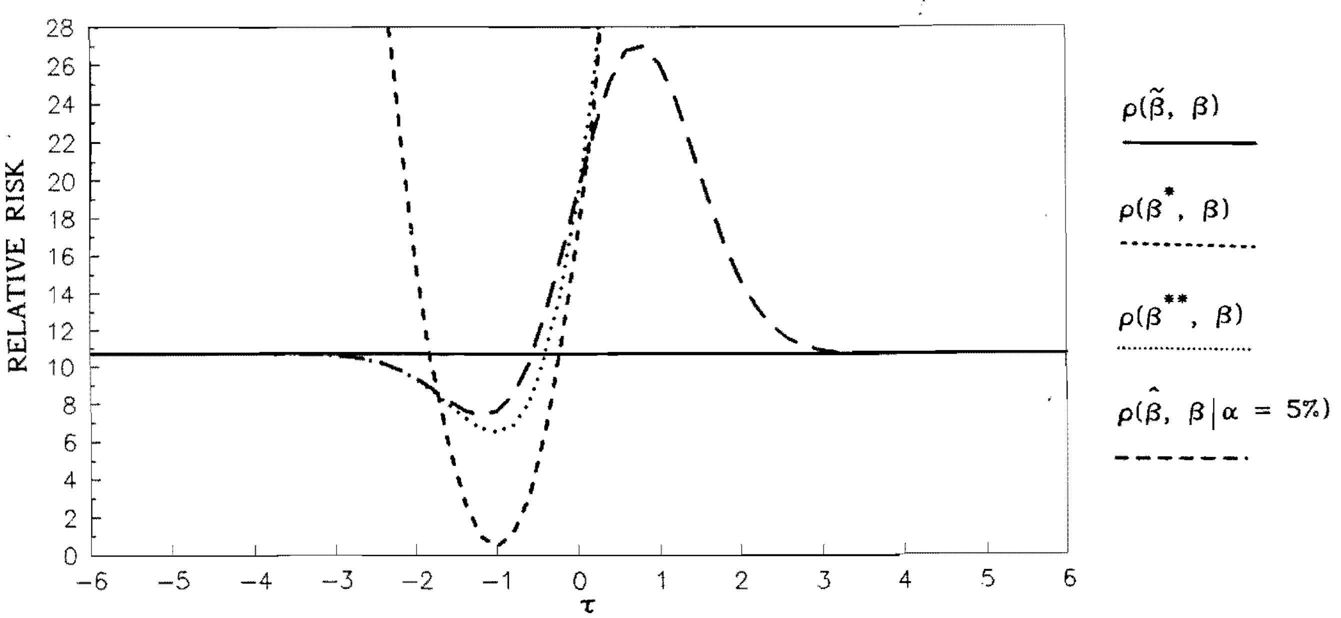


Figure 4.12

Relative risk functions of $\tilde{\beta}$, β^{*}, β^{**} and $\hat{\beta}$ in a mis-specified model

income series from spirits data : n = 60
omitted regressors : price series from spirits data (β_3 = -0.5)



CHAPTER FIVE

AN OPTIMAL CRITICAL VALUE OF A PRE-TEST FOR AN INEQUALITY RESTRICTION WHEN ESTIMATING THE PREDICTION VECTOR

5.1 INTRODUCTION

In Chapter 4, we derived the predictive risk functions of the inequality restricted and pre-test estimators in a linear model which is underfitted. The risks of these estimators in a properly specified model, as derived by Judge and Yancey (1981, 1986), are also embodied in our results as a special case. We note from our results in the preceding chapter that if the researcher is to decide on whether to place an inequality restriction on the coefficients, then for any given level of model specification error, his best strategy would be to use either the unrestricted or the inequality restricted estimators according to the degree of constraint specification error. For any given λ_1, pre-testing is never preferred, as the risk of the inequality pre-test estimator is always no less than the minimum of the risks of its component estimators. However, because λ_1 is unknown in practice, a pre-test procedure is still routinely used by many applied researchers.

In most applied situations, the traditional 1% or 5% significance levels, are often chosen for the pre-test, without any necessary theoretical justification. This arbitrariness of the choice of the significance level can often lead to undesirable risk properties of the resulting pre-test estimators, as is evidenced in other pre-test problems that have received attention in the literature (see, for example, Brook (1976)). To avoid this problem, various explicit criteria have been proposed to set optimal levels of significance for pre-tests in various contexts. For instance, in the case of pre-testing of linear equality constraints on regression coefficients, Sawa and Hiromatsu (1973), Brook (1976) and Toyoda and Wallace (1976) obtained optimal critical values when estimating the prediction vector according to the criteria of mini-max regret and minimizing the average relative risk. These criteria were also used to select an optimal significance level for a preliminary test of variance homogeneity (Toyoda and Wallace (1975) and Ohtani and Toyoda (1978)). Other criteria, such as that based on an unbiased decision rule (Ohtani (1992)) and Bayesian minimum expected loss (Bock *et al.* (1973a)) have also been proposed in the literature for choosing the optimal pre-test size in other contexts.

The purpose of this chapter is to address the problem of suggesting optimal critical values for the preliminary test of an inequality restriction when estimating E(y), using both the criteria of mini-max regret and minimizing the average relative risk. This is considered within the context of an underfitted model, which encompasses the properly specified model as a special case. Although the predictive risk functions of the inequality pre-test estimator were derived in the preceding chapter for both the σ^2 known and unknown cases, our discussion of an optimal critical value will only proceed for the latter case, as typically σ^2 is unknown.

5.2 THE PROBLEM REVISITED

As in Chapter 4, the statistical model under consideration is

$$y = X\beta + Z\eta + \varepsilon \qquad ; \qquad \varepsilon \sim N(0, \sigma^2 I) \tag{5.1}$$

where y, X, β, Z, η and ε are defined as previously. The inequality restriction to be tested is represented by the hypothesis H_0: $C'\beta \geq r$ *vs* H_1: $C'\beta < r$, with $\bar{\tau} = r - C'\beta$. It is assumed that in specifying the model, the researcher mistakenly omits the set of regressors Z, and so the fitted model is

$$y = X\beta + \mu \qquad ; \qquad \mu \sim N(Z\eta,\ \sigma^2 I). \tag{5.2}$$

H_0 is tested using the statistic $t'' = (C'\tilde{\beta} - r)/s.e(C'\tilde{\beta})$. t'' has a non-central t distribution with v degrees of freedom and non-centrality parameters $\lambda_1^2 = (\bar{\tau} - C'S^{-1}X'Z\eta)'(C'S^{-1}C)^{-1}(\bar{\tau} - C'S^{-1}X'Z\eta)/(2\sigma^2)$ and $\lambda_2 = \eta'Z'(I-XS^{-1}X')Z\eta/(2\sigma^2)$. However, as the researcher is unaware of the specification error in the model, he believes that t'' has a central t distribution when $C'\beta = r$. Hence the critical value c is found by solving

$$\int_{-\infty}^{c} dt_v = \alpha \tag{5.3}$$

for a given level of significance, α.

The use of the preliminary test leads to the inequality pre-test estimator (IPTE), $\hat{\beta}$, which is the unrestricted estimator (UE), $\tilde{\beta}$, if H_0 is rejected, and the inequality restricted estimator (IRE), β^{**}, if we cannot reject H_0. Rather like a pre-test estimator, β^{**} is itself a choice between the equality restricted estimator (ERE), β^{*}, and the unrestricted estimator (UE), $\tilde{\beta}$, depending on whether the restriction is binding.

The predictive risk functions of these estimators were derived and illustrated in Chapter 4. Those results show that neither the IPTE nor its component estimators strictly dominate one another. For any given value of λ_2, the IPTE is never the minimum risk estimator and, over some regions in the λ_1 space, it is the estimator with the highest risk. From the point of view of minimizing an estimator's risk, it would seem to be desirable to have a rule that mixes the unrestricted and restricted estimators according to the magnitude of λ_1. However, given that λ_1 is unknown, the choice of estimators is unclear in practice, and we would like to select an optimal critical value that brings the risk of the IPTE as close as possible to the smallest that it can achieve. The particular critical value chosen depends, as we would expect, on the optimality criterion adopted.

5.3 A BROOK-TYPE MINI-MAX REGRET CRITERION

One of the criteria that has been suggested in the literature for determining an optimal critical value in other pre-test contexts is that of mini-max regret. Along the lines of Sawa and Hiromatsu (1973) and Brook (1976), we define the regret function as

$$\mathrm{REG}(\lambda_1,\ c) = \rho(X\hat{\beta},\ E(y)) - \inf_{c} \rho(X\hat{\beta},\ E(y)), \tag{5.4}$$

where $\inf_{c} \rho(X\hat{\beta},\ E(y))$ is the infimum (which equals the minimum in this context) of $\rho(X\hat{\beta},\ E(y))$ over all values of λ_1, and is therefore the minimum risk boundary of the pre-test predictor. As mentioned in Chapter 4, for any particular value of λ_2, it can be shown that[1]

$$\inf_{c} \rho(X\hat{\beta},\ E(y)) = \begin{cases} \rho(X\hat{\beta},\ E(y)\,|\,c = -\infty) = \rho(X\beta^{**},\ E(y)) \text{ if } \lambda_1 < \lambda_1^* \\ \\ \rho(X\hat{\beta},\ E(y)\,|\,c = 0) \quad = \rho(X\tilde{\beta},\ E(y)) \quad \text{ if } \lambda_1 \geq \lambda_1^* \end{cases}, \tag{5.5}$$

where λ_1^* is the point at which one would switch from using $X\beta^{**}$ to $X\tilde{\beta}$. The minimum risk boundary of the inequality pre-test predictor is indicated by the upper boundary of the shaded area in Figure 5.1 (see p. 134). From the definition of (5.4), regret, for any given levels of λ_1 and c, is the difference between the risk of the pre-test estimator associated with that particular critical value and the minimum possible pre-test estimator risk across all critical values. For $\lambda_1 < \lambda_1^*$, $\mathrm{Reg}(\lambda_1, c)$ takes on a maximum at λ_1^L, and this value of regret is labelled d^L in Figure 5.2 (see p. 135). For $\lambda_1 >$

---

[1] The proof for (5.5) follows the same approach for proving the infimum of $\hat{\sigma}^2_{ML}$ given in Appendix 7A. Full details are available upn request.

λ_1^*, the maximum of the regret function, denoted by d^U, occurs at λ_1^U. For any given c, d^L and d^U are therefore the maximum additional penalty for choosing that particular level of critical value instead of $c = -\infty$ (for $\lambda_1 \le \lambda_1^*$) and $c = 0$ (for $\lambda_1 > \lambda_1^*$) respectively.

Analogous to the case in which the linear restriction holds as a strict equality, the mini-max regret procedure is to seek a critical value which makes both d_L and d_U as small as possible. However, it is found empirically that increasing $|c|$ decreases d_L but increases d_U (see Figure 5.2). Because of this monotonicity property, to minimize the maximum regret over all values of c and λ_1 amounts to finding $c = c^{MX}$ such that d_L and d_U are equal, or in other words,

$$\sup_{\lambda_1<\lambda_1^*} REG(\lambda_1, c^{MX}) = \sup_{\lambda_1\ge\lambda_1^*} REG(\lambda_1, c^{MX}), \tag{5.6}$$

where c^{MX} is the mini-max regret critical value.

As it seems impossible to derive the mini-max regret critical values analytically, we rely on numerical computations. Brent's (1974) algorithm is used to search for the value of λ_1^*. It is found that regardless of the magnitude of λ_2, λ_1^* is approximately[2] 0.594. The Golden Section Search Routine given in Press *et al.* (1986) is used to compute the mini-max regret critical values. These routines are incorporated into a FORTRAN program which has been executed on a VAX 6340 machine. The mini-max regret critical values, the corresponding level of α, and the least favourable values, λ_1^L and λ_1^U, of λ_1 and are presented in Table 5.1 (pp. 128-130) for several values of λ_2.

Table 5.1 shows that when the model is properly specified (*i.e.*, $\lambda_2 = 0$), the optimal critical value fluctates only slightly with the model's degrees of

---

[2] When there is no mis-specification in model, this result implies that the risk of β^{**} crosses that of $\tilde{\beta}$ at $\tau/\sigma = \sqrt{2} \times 0.594 = 0.84$. This finding is previously noted by Thomson and Schmidt (1982).

freedom. For moderate to high degrees of freedom, the optimal critical value does not vary much from -1.12. Although the optimal critical value is approximately invariant to changes in the model's degrees of freedom, the size of the test which is associated with the optimal critical value is not. The optimal size decreases as the degrees of freedom increase. However, it is apparent that for small to moderate degrees of freedom, the optimal size is greater than the traditional 5% level. Clearly, the use of the optimal pre-test size, as compared to the 5% level of significance, will increase the probability of the unrestricted estimator being chosen. This has the effect of bringing the risk of the inequality pre-test predictor closer to that of the unrestricted predictor, as shown in Figures 5.3 and 5.4 (p. 136). These results are qualitatively consistent with the those of Sawa and Hiromatsu (1973) and Brook (1976) for the case in which the linear restriction holds as a strict equality.[3]

Once we allow for the omission of relevant regressors from the model (*i.e.*, $\lambda_2 > 0$), the optimal critical values vary with the degrees of freedom of the model and can differ considerably from -1.12. This feature is also shown in Table 5.1. Accordingly, any attempt to apply a mini-max regret critical value obtained under the assumption that $\lambda_2 = 0$ will not necessarily lead to an optimal pre-test risk when the model is underfitted (see Figures 5.5 - 5.8 for example on pp. 137-138). The rate at which c^{MX} changes with the degrees of freedom in the model increases as λ_2 increases. Furthermore, for any given v, $|c^{MX}|$ decreases as λ_2 increases. Accordingly, the optimal pre-test size increases monotonically with λ_2, for given degrees of freedom. Again, these

---

[3] Brook (1976) also tabulates mini-max regret critical values when there is more than one equality restriction, and finds that generally, the results are qualitatively similar to those obtained when there is only a single restriction.

results are qualitatively consistent with those reported in the literature for the case in which the linear restriction is held as a strict equality and the model is mis-specified through the omission of relevant regressors (see Giles *et al.* (1992a)).

Table 5.1 also shows that for any fixed λ_2, as the model's degrees of freedom increases, λ_1^L increases but λ_1^U decreases. However, the degree of variation is typically very slight. Again, this is consistent with the result of Brook (1976) for the case in which the prior restriction on β exists in the form of a linear equality restriction. Finally, both the lower and upper least favourable values of λ_1 are roughly constant across different values of λ_2, reflecting the fact that for any given λ_1, the predictive risk of $\hat{\beta}$ in an underfitted model has essentially the same characteristics as in a properly specified model.

5.4 AN ALTERNATIVE MINI-MAX REGRET CRITERION

The discussion above defines regret as the difference between the risk of the inequality pre-test estimator for a particular value of c and the minimum possible pre-test estimator risk across all values of c and λ_1. The procedure then seeks to select a critical value that minimizes the maximum value of that regret. An alternative mini-max regret criterion is to define the regret function as

$$\overline{\mathrm{Reg}}(\lambda_1, c) = \rho(X\hat{\beta}, E(y)) - \min[\rho(X\beta^*, E(y)), \inf_c \rho(X\hat{\beta}, E(y))]. \qquad (5.7)$$

This criterion takes into account the equality restricted predictor, which does not belong to the family of inequality pre-test estimators,[4] but has risk lower

[4] This is because the subsequent step of checking whether $C'\tilde{\beta} \geq r$ is always assumed if the null hypothesis is accepted, hence although explicitly the

than $\inf_c \rho(X\hat{\beta}, E(y))$ over certain regions in the λ_1 space. This typically occurs when $|\lambda_1|$ is sufficiently small. It can be seen from Figure 5.9 that for $\lambda_1 \le \bar{\lambda}_1^*$, imposing the inequality constraint is the preferred strategy. When $\bar{\lambda}_1^* \le \lambda_1 \le \bar{\lambda}_1^{**}$, the best strategy is to impose the exact restriction $C'\beta = r$. Finally, when $\lambda_1 > \bar{\lambda}_1^{**}$, it is best to ignore the prior information. Therefore,

$$\min[\rho(X\beta^*, E(y)), \inf_c \rho(X\hat{\beta}, E(y))]$$

$$= \begin{cases} \rho(X\beta^{**}, E(y)) = \rho(X\hat{\beta}, E(y)|c = -\infty) & \text{if } \lambda_1 \le \bar{\lambda}_1^* \\ \rho(X\beta^*, E(y)) & \text{if } \bar{\lambda}_1^* \le \lambda_1 \le \bar{\lambda}_1^{**} \\ \rho(X\tilde{\beta}, E(y)) = \rho(X\hat{\beta}, E(y)|c = 0) & \text{if } \lambda_1 > \bar{\lambda}_1^{**} \end{cases} \quad . \qquad (5.8)$$

Let $\bar{d}_{L1}$, $\bar{d}_{L2}$ and $\bar{d}_U$ denote the respective maximum distances associated with the regret function in regions A, B and C given in Figure 5.9 (see p. 139). It is found numerically that increasing $|c|$ decreases $\bar{d}_{L1}$ and $\bar{d}_{L2}$, but increases $\bar{d}_U$, and that $\bar{d}_{L1} < \bar{d}_{L2}$ for all values of c. Therefore, the mini-max regret critical value, under this alternative definition of the regret function, denoted by c^{MX^*}, is the value of c which equalizes the distance labelled $\bar{d}_{L2}$ and $\bar{d}_U$ in Figure 5.9.

Table 5.2 (see p. 131) presents the results when mini-max regret critical values are calculated according to this alternative definition of the regret function. It shows that the use of (5.7) as the regret function generally reduces the optimal critical values and so results in a more frequent acceptance of the null, but qualitatively, the pattern of the results is the same as is reported in Table 5.1. In particular, c^{MX*} is roughly contant when

---

pre-test estimator may take the form of the equality restricted estimator, implicitly it is the inequality restricted estimator that is being used. This is reflected in the fact that there is no value of c such that the inequality pre-test estimator degenerates to the equality restricted estimator with probability one.

$\lambda_2 = 0$, but varies with the model's degrees of freedom when $\lambda_2 > 0$.

Figures 5.10 and 5.11 (see p. 140) depict the predictive risk functions of the inequality pre-test estimators corresponding to $c = c^{MX}$ and $c = c^{MX*}$ for the cases of $n = 40$, $k = 5$, $\lambda_2 = 0$ and $n = 60$, $k = 10$, $\lambda_2 = 10$ respectively. We see that neither estimator uniformly dominates the other. This result holds for all values of λ_2 that we have considered.

5.5 THE CRITERION OF MINIMIZING THE AVERAGE RELATIVE RISK

Another possible criterion for determining the optimal critical value is that of minimizing the average relative risk over the entire range of critical values for rejecting or accepting H_0. This is equivalent to minimizing the area between the pre-test risk function and the minimum risk boundary. This is the approach used by Toyoda and Wallace (1976), who derive and present optimal critical values for a pre-test of exact linear restrictions when estimating E(y) in the standard linear model. In the context of our problem, the area between the pre-test risk function and the minimum risk boundary can be expressed as

$$A(c) = \int_{-\infty}^{\infty} [\rho(X\hat{\beta},\ E(y)) - \inf_{c} \rho(X\hat{\beta},\ E(y))]\ d\lambda_1 \quad . \tag{5.9}$$

As opposed to minimizing the maximum regret, this procedure chooses a critical value that minimizes the "average" regret of not being on the minimum risk boundary[5]. Now, from the preceding chapter, we know that, if $c < 0$, then

[5] As $\inf_c \rho(X\hat{\beta},\ E(y))$ is equivalent to $\min\ [\rho(X\tilde{\beta},\ E(y)),\ \rho(X\beta^{*},\ E(y))]$, which does not depend on c, minimizing A(c) is, in a sense, equivalent to minimizing the area under the risk function of $X\hat{\beta}$. This also suggests that the use of the alternative definition of minimum boundary as $\min\ [\rho(X\beta^{**},\ E(y)),\ \inf_c \rho(X\hat{\beta},\ E(y))]$ will not affect the result obtained. However, as noted by Toyoda and

$$\rho(X\beta,\ E(y)) = \begin{cases} k + 2\lambda_2 + (E_{3,v} - P_3)/2 - \lambda_1^2(E_{1,v} - P_1) & \text{if } \lambda_1 \le 0 \\ k + 2\lambda_2 - 1 + 2\lambda_1^2 + (E_{3,v} + P_3)/2 & , \\ \quad - \lambda_1^2(E_{1,v} + P_1) - 2\lambda_1^2 G_{1,v} + G_{3,v} & \text{if } \lambda_1 > 0 \end{cases} \tag{5.10}$$

where

$$P_I = P(\chi_I^2 \ge 2\lambda_1^2),$$

$$E_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)} \int_0^{\infty} P(\chi_I^2 \ge (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2)(q_J^2)^{J/2+t-1} e^{-q_J^2/2}\, dq_J^2,$$

$$G_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)} \int_0^{2v\lambda_1^2/c^2} P(\chi_I^2 < (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2)(q_J^2)^{J/2+t-1} e^{-q_J^2/2}\, dq_J^2,$$

$I = 1,\ 3$ and q_J^2 is a non-central chi-square random variable with J degrees of freedom and non-centrality parameter λ_2 ; *i.e.* $q_J^2 \sim \chi'^2_{(J;\lambda_2)}$.

Alternatively, if $c \ge 0$, then

$$\rho(X\hat{\beta},\ E(y)) = \rho(X\tilde{\beta},\ E(y)) = k + 2\lambda_2. \tag{5.11}$$

Using these results and making some necessary substitutions, we obtain the following :

if $c < 0$, then

$$A(c) = \int_{-\infty}^{\infty} (E_{3,v}/2 - \lambda_1^2 E_{1,v})\, d\lambda_1 + \int_0^{\infty} (G_{3,v} - 2\lambda_1^2 G_{1,v})\, d\lambda_1$$
$$+ \int_{\lambda_1^*}^{\infty} (2\lambda_1^2 - 1 + P_3/2 - \lambda_1^2 P_1)\, d\lambda_1 \quad , \tag{5.12}$$

if $c \ge 0$, then

Wallace (1976), subtracting the shaded area has the convenient virtue of making the resultant integral bounded.

$$A(c) = \int_{-\infty}^{0} (P_3/2 - \lambda_1^2 P_1)\, d\lambda_1 - \int_{0}^{\lambda_1^*} (P_3/2 + 2\lambda_1^2 - \lambda_1^2 P_1 - 1)\, d\lambda_1, \tag{5.13}$$

which is a constant with respect to c.

When $c \to 0^-$, $E_{1,v} \to P_1$, $E_{3,v} \to P_3$, $G_{3,v} \to 1 - P_3$ and $G_{1,v} \to 1 - P_1$. Accordingly,

$$\lim_{c\to 0^-} A(c) = \int_{-\infty}^{0} (P_3/2 - \lambda_1^2 P_1)\, d\lambda_1 - \int_{0}^{\lambda_1^*} (P_3/2 + 2\lambda_1^2 - \lambda_1^2 P_1 - 1)\, d\lambda_1 = A(0).$$

A(c) is therefore continuous at c = 0.

It is obvious that when $c \geq 0$, $\frac{\partial A}{\partial c} = 0$. Now k, λ_2, P_1 and P_3 are independent of c, hence when $c < 0$,

$$\frac{\partial A}{\partial c} = \frac{\partial}{\partial c}\left\{ \int_{-\infty}^{\infty}\left[\frac{E_{3,v}}{2} - \lambda_1^2 E_{1,v}\right] d\lambda_1 + \int_{0}^{\infty} \left[G_{3,v} - 2\lambda_1^2 G_{1,v}\right] d\lambda_1\right\}$$

$$= \frac{\partial}{\partial c}\, e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{1/2+t}\Gamma\left(\frac{v}{2}+t\right)} \left\{ \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{d^2}^{\infty} \left(\frac{u_1^{1/2} e^{-u_1/2}}{2\, 2^{3/2}\Gamma(\frac{3}{2})} - \frac{\lambda_1^2 u_1^{-1/2} e^{-u_1/2}}{\sqrt{2}\Gamma(\frac{1}{2})}\right)\right.$$

$$\left. + \int_{0}^{\infty}\int_{0}^{2v\lambda_1^2/c^2}\int_{0}^{d^2} \left(\frac{u_1^{1/2} e^{-u_1/2}}{2^{3/2}\Gamma(\frac{3}{2})} - \frac{2\lambda_1^2 u_1^{-1/2} e^{-u_1/2}}{\sqrt{2}\Gamma(\frac{1}{2})}\right) \right\} (q_v^2)^{v/2+t-1}\, e^{-q_v^2/2}\, du_1\, dq_v^2\, d\lambda_1$$

where $d = cq_v/\sqrt{v} + \sqrt{2}\lambda_1$. (5.14)

Now recognising that $\frac{\partial}{\partial c}\int_{d^2}^{\infty}\left(\ \cdot\ \right) = -\frac{\partial}{\partial c}\int_{0}^{d^2}\left(\ \cdot\ \right)$, and that $2^{3/2}\Gamma(\frac{3}{2}) = \sqrt{2}\Gamma(\frac{1}{2}) = \sqrt{2\pi}$, we can write (5.14) as

$$\frac{\partial A}{\partial c} = \frac{\partial}{\partial c}\, e^{-\lambda_2} \sum_{t=0}^{\infty} \frac{\lambda_2^t}{t!\,2^{1/2+t}\Gamma\left(\frac{v}{2}+t\right)} \left\{ \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{0}^{d^2} \left(\frac{\lambda_1^2 u_1^{-1/2} e^{-u_1/2}}{\sqrt{2\pi}} - \frac{u_1^{1/2} e^{-u_1/2}}{2\sqrt{2\pi}} \right)\right.$$

$$+\int_0^{\infty}\int_0^{2v\lambda_1^2/c^2}\int_0^{d^2}\left(\frac{u_1^{1/2}e^{-u_1/2}}{\sqrt{2\pi}} - \frac{2\lambda_1^2 u_1^{-1/2}e^{-u_1/2}}{\sqrt{2\pi}}\right)\Bigg\}(q_v^2)^{v/2+t-1}e^{-q_v^2/2}\,du_1\,dq_v^2\,d\lambda_1$$

$$= e^{-\lambda_2}\sum_{t=0}^{\infty}\frac{\lambda_2^t}{t!\,2^{1/2+t}\Gamma\left(\frac{v}{2}+t\right)}\int_{-\infty}^{\infty}\int_0^{\infty}\frac{(2\lambda_1^2 - d^2)dq_v e^{-d^2/2}}{\sqrt{d^2}\sqrt{2\pi v}}(q_v^2)^{v/2+t-1}e^{-q_v^2/2}dq_v^2\,d\lambda_1$$

$$+ e^{-\lambda_2}\sum_{t=0}^{\infty}\frac{\lambda_2^t}{t!\,2^{1/2+t}\Gamma\left(\frac{v}{2}+t\right)}\int_0^{\infty}\left\{\int_0^{2v\lambda_1^2/c^2}\frac{(d^2 - 2\lambda_1^2)2q_v e^{-d^2/2}}{\sqrt{2\pi v}}(q_v^2)^{v/2+t-1}e^{-q_v^2/2}dq_v^2\right.$$

$$\left.+\left[P\left(\chi_3^2 < (-\sqrt{2}\lambda_1 + \sqrt{2}\lambda_1)^2\right) - 2\lambda_1^2 P\left(\chi_1^2 < (-\sqrt{2}\lambda_1 + \sqrt{2}\lambda_1)^2\right)\right](-4v\lambda_1^2/c^3)\right\}d\lambda_1^2$$

$$= e^{-\lambda_2}\sum_{t=0}^{\infty}\frac{\lambda_2^t}{t!\,2^{1/2+t}\Gamma\left(\frac{v}{2}+t\right)}\left\{\int_{-\infty}^{\infty}\int_0^{\infty} I(d) - \int_0^{\infty}\int_0^{2v\lambda_1^2/c^2} 2\right\}$$

$$\times\ \frac{(2\lambda_1^2 - d^2)q_v e^{-d^2/2}}{\sqrt{2\pi v}}(q_v^2)^{v/2+t-1}e^{-q_v^2/2}dq_v^2\,d\lambda_1 , \qquad (5.15)$$

where I(d) takes the value of 1 if $d > 0$, -1 otherwise.[6] It can be shown that A(c) approaches zero if (i) $c \to -\infty$, or (ii) $c \to 0$.

So, given that A(c) is continuous at $c = 0$, and $\frac{\partial A}{\partial c} \to 0$ as $c \to 0^-$, A(c) therefore attains a stationary point at $c = 0$ if and only if A(c) is also

---

[6] This is because given that $d = cq_j/\sqrt{v} + \sqrt{2}\lambda_1$, when λ_1 is unrestricted with respect to sign (as in the first part of (5.14)), the sign of d is also undetermined. If d is negative, then $\sqrt{d^2} = -d$, else $\sqrt{d^2} = d$. However, if λ_1 is known to be positive and the values of q_j are restricted to the range of $[0, -\sqrt{2v}\lambda_1/c]$ (as in the second part of (5.14)), then d must be non-negative.

differentiable at $c = 0$.[7] The differentiability of $A(c)$ at $c = 0$ is proven in Appendix 5B. It then follows that $A(c)$ attains a stationary point at $c \doteq 0$.

It is clear that $c = -\infty$ does not yield a minimum since $A(-\infty)$ does not converge. To ensure that $c = 0$ is a minimum, we have performed numerical computations of $A(c)$ using the trapezoidal rule for various values of λ_2 and v at equally spaced intervals of 0.02 for $\lambda_1 = [-10, 10]$ (*i.e.* 1000 intervals), beginning from $c = -10$, and giving increments of 0.1 for each case. Figures 5.12 and 5.13 (see p. 141) illustrate some of the results. They show that for given values of v and λ_2, $A(c)$ is maximized when c is sufficently small, decreases in the interval $(-\infty, 0)$ before reaching a minimum at $c = 0$. However, since $A(c)$ is continuous at $c = 0$ and is constant with respect to c for $c \geq 0$, it follows that every value of c in the subset $\{c | c \in R^+\}$ also yields a minimum. So the minimum is not unique. This does not introduce any difficulty regarding the applicability of our result, but merely indicates that the average relative risk is minimized when the unrestricted estimator is used.

5.6 CONCLUDING DISCUSSION

In this chapter we have considered the problem of determining the critical value of a prior test of an inequality restriction using two commonly adopted optimality criteria given in the literature. Consistent with the results of Toyoda and Wallace (1976) for the case in which the prior restriction on β exists in the form of a single linear equality and the underlying model is

---

[7] Arguably, given that $A(c)$ is constant for $c \geq 0$ and continuous at $c = 0$, $c = 0$ must be a stationary point if $A(c)$ is also differentiable at $c = 0$. Therefore, it is not necessary to differentiate $A(c)$ to prove that $c = 0$ is a stationary point. However, differentiation is necessary if we want to check whether there exist any other stationary points.

properly specified, the pre-test critical value that we derive according to the criterion of minimum average relative risk leads to the choice of the unrestricted estimator, regardless of the model's degrees of freedom. Our finding is also invariant to the extent to which the model is mis-specified. Given Toyoda and Wallace's results, it is unclear whether our findings would extend to more than one restriction.[8]

If an alternative mini-max regret criterion is used, then it is found that under the maintained assumption of a properly specified model, the optimal critical value is approximately -1.12 regardless of the model's degrees of freedom. However, this property no longer holds once we allow for possible mis-specification in the model. Accordingly, any attempt to apply an optimal critical value obtained under the assumption that $\lambda_2 = 0$ will not necessarily lead to an optimal pre-test risk when the model is in fact underfitted.

While these two criteria lead to different choices of optimal critical values, the one based on the criterion of minimizing the average relative risk has an obvious practical appeal as the result obtained is independent of the degree of specification error in the model. By constrast, the mini-max regret approach suffers from the disadvantage of being sensitive to the extent to which the model is mis-specified. However, no matter which criterion we adopt, in general it is apparent that the optimal pre-test size is frequently much greater than the traditionally used one or five precent significance levels.

Having said this, the practicality of our results is still limited even when the model is correctly specified, as they are based on the risk function

---

[8] Toyoda and Wallace (1976) find that when there are more than 5 equality restrictions, the optimal critical value increases with both the degrees of freedom and the number of restrictions, and is approximately 2 for the central F distribution when the number of constraints is more than 60.

of the inequality pre-test predictor, which is equivalent to the risk of the estimator for the coefficient vector only if the columns in the regressor matrix are orthonormal. In practice, the data are likely to be collinear. Brook and Fletcher (1981) show that when pre-testing for linear equalities, optimal critical values obtained under the assumption of orthonormal data are not optimal when the data exhibits high multicollinearity. Given their results, it is unclear whether our findings will carry over to the non-orthonormal case. The extension of our analysis to the non-orthonormal model remains an interesting topic for further research.

APPENDIX 5A

Table 5.1 : Optimal critical values according to the Brook-type mini-max regret criterion, the corresponding percentage levels of the t test, and least favourable values of λ_1.

λ_2	v	c^*	$\alpha(\%)$	λ_1^L	λ_1^U
0	2	-1.112	19.096	0.175	1.304
	5	-1.118	15.717	0.210	1.213
	10	-1.122	14.395	0.224	1.174
	15	-1.124	13.929	0.230	1.167
	20	-1.125	13.691	0.233	1.161
	25	-1.126	13.547	0.234	1.157
	30	-1.126	13.449	0.235	1.154
	35	-1.127	13.380	0.236	1.152
	40	-1.127	13.327	0.237	1.151
	45	-1.127	13.286	0.237	1.150
	50	-1.127	13.253	0.238	1.150
	55	-1.127	13.227	0.238	1.149
	60	-1.127	13.204	0.238	1.148
	65	-1.127	13.185	0.239	1.147
	70	-1.128	13.169	0.239	1.147
	75	-1.128	13.155	0.239	1.147
	80	-1.128	13.142	0.239	1.147
2	2	-0.642	29.322	0.198	1.239
	5	-0.834	22.115	0.215	1.200
	10	-0.949	18.253	0.226	1.175
	15	-0.999	16.682	0.230	1.166
	20	-1.027	15.829	0.233	1.161
	25	-1.045	15.293	0.234	1.157
	30	-1.058	14.926	0.236	1.154
	35	-1.067	14.658	0.236	1.153
	40	-1.074	14.455	0.237	1.152
	45	-1.080	14.295	0.237	1.151
	50	-1.085	14.165	0.238	1.149
	55	-1.088	14.059	0.238	1.149
	60	-1.092	13.970	0.238	1.148
	65	-1.094	13.894	0.239	1.148
	70	-1.097	13.828	0.239	1.147
	75	-1.099	13.771	0.239	1.147
	80	-1.100	13.722	0.239	1.147
10	2	-0.338	38.363	0.226	1.174
	5	-0.502	31.834	0.229	1.169
	10	-0.649	26.538	0.232	1.162
	15	-0.737	23.632	0.233	1.159

(Table 5.1 (cont'd))

λ_2	v	c^*	$\alpha(\%)$	λ_1^L	λ_1^U
10	20	-0.796	21.765	0.235	1.156
	25	-0.839	20.458	0.236	1.154
	30	-0.873	19.490	0.236	1.152
	35	-0.899	18.705	0.237	1.151
	40	-0.920	18.150	0.237	1.150
	45	-0.938	17.667	0.238	1.150
	50	-0.953	17.267	0.238	1.149
	55	-0.965	16.929	0.238	1.148
	60	-0.976	16.640	0.239	1.148
	65	-0.986	16.390	0.239	1.148
	70	-0.994	16.172	0.239	1.147
	75	-1.002	15.980	0.239	1.147
	80	-1.009	15.810	0.239	1.146
25	2	-0.221	42.287	0.235	1.156
	5	-0.340	37.401	0.235	1.155
	10	-0.460	32.774	0.236	1.153
	15	-0.541	29.816	0.237	1.152
	20	-0.602	27.687	0.237	1.151
	25	-0.651	26.061	0.238	1.150
	30	-0.690	24.769	0.238	1.150
	35	-0.723	23.715	0.238	1.150
	40	-0.752	22.835	0.238	1.149
	45	-0.776	22.093	0.239	1.148
	50	-0.797	21.454	0.239	1.147
	55	-0.816	20.900	0.239	1.147
	60	-0.833	20.414	0.239	1.147
	65	-0.848	19.984	0.239	1.147
	70	-0.861	19.601	0.239	1.146
	75	-0.874	19.258	0.239	1.146
	80	-0.885	18.949	0.240	1.146
50	2	-0.158	44.455	0.238	1.150
	5	-0.246	40.774	0.238	1.149
	10	-0.340	37.049	0.238	1.149
	15	-0.407	34.482	0.238	1.148
	20	-0.460	32.514	0.239	1.148
	25	-0.504	30.916	0.239	1.147
	30	-0.542	29.603	0.239	1.147
	35	-0.574	28.477	0.239	1.147
	40	-0.603	27.504	0.239	1.146
	45	-0.628	26.652	0.239	1.146

(Table 5.1 (cont'd))

λ_2	v	c^*	$\alpha(\%)$	λ_1^L	λ_1^U
50	50	-0.660	25.617	0.239	1.146
	55	-0.672	25.227	0.240	1.146
	60	-0.691	24.623	0.240	1.146
	65	-0.708	24.078	0.240	1.146
	70	-0.724	23.582	0.240	1.145
	75	-0.738	23.130	0.240	1.145
	80	-0.752	22.715	0.240	1.145

Table 5.2 : Optimal critical values according to the alternative mini-max regret criterion, the corresponding percentage levels of the t test, and least favourable values of λ_1.

λ_2	v	c^*	$\alpha(\%)$	λ_1^L	λ_1^U
0	5	-1.499	9.713	0.144	1.409
	15	-1.499	7.729	0.171	1.338
	35	-1.501	7.110	0.175	1.328
	50	-1.503	6.960	0.184	1.311
	80	-1.513	6.707	0.186	1.306
2	5	-1.125	15.588	0.186	1.722
	15	-1.335	10.090	0.213	1.655
	35	-1.421	8.211	0.225	1.630
	50	-1.443	7.769	0.227	1.624
	80	-1.462	7.376	0.230	1.618
10	5	-0.672	26.570	0.210	1.662
	15	-0.983	17.070	0.219	1.642
	35	-1.196	11.983	0.226	1.628
	50	-1.267	10.553	0.228	1.623
	80	-1.340	9.196	0.230	1.618
25	5	-0.452	33.500	0.222	1.635
	15	-0.720	24.119	0.225	1.629
	35	-0.962	17.135	0.228	1.623
	50	-1.060	14.717	0.229	1.620
	80	-1.175	12.164	0.231	1.617
50	5	-0.341	37.348	0.228	1.630
	15	-0.541	29.812	0.229	1.621
	35	-0.763	22.527	0.230	1.619
	50	-0.865	19.555	0.231	1.618
	80	-0.999	16.044	0.231	1.616

APPENDIX 5B

Lemma 5.1 :

$\partial A(c)/\partial c$ approaches zero if $c \to -\infty$ or $c \to 0$.

Proof :

Recall that $d = cq_v/\sqrt{v} + \sqrt{2}\lambda_1$. Hence $d \to -\infty$ as $c \to -\infty$, and $d \to \sqrt{2}\lambda_1$ as $c \to 0$ for any given λ_1 and q. Now, $\lim_{c \to -\infty} (2\lambda_1^2 - d^2)e^{-d^2/2} = 0 - \lim_{c \to -\infty} d^2/e^{d^2/2}$. Note that $\lim_{c \to -\infty} d^2/e^{d^2/2}$ takes the indeterminate form of ∞/∞. Applying L' Hospital's rule, $\lim_{c \to -\infty} d^2/e^{d^2/2} = \lim_{c \to -\infty} 2d/(de^{d^2/2}) = \lim_{c \to -\infty} 2/e^{d^2/2} = 0$. Hence $\frac{\partial A}{\partial c} \to 0$ as $c \to -\infty$. Similarly, $\lim_{c \to 0^-} (2\lambda_1^2 - d^2) = \lim_{c \to 0^-} (2\lambda_1^2 - (cq_v + \sqrt{2}\lambda_1)^2) = 0$. Therefore, $\frac{\partial A}{\partial c} \to 0$ as $c \to 0^-$. Q.E.D.

Lemma 5.2 :

A(c) is differentiable at c = 0.

Proof :

To prove that A(c) is differentiable at c = 0, we must show that $\lim_{c \to 0} \frac{A(c) - A(0)}{c - 0}$ exists.[9] Now it is clear that

$$\lim_{c \to 0^+} \frac{A(c) - A(0)}{c - 0} = \lim_{c \to 0^+} \frac{0}{c - 0} = \lim_{c \to 0^+} 0 = 0.$$

Unfortunately, $\lim_{c \to 0^-} \frac{A(c) - A(0)}{c - 0}$ takes the indeterminate form of 0/0. The evaluation of this limit therefore necessitates the use of L' Hospital's rule. Applying L' Hospital's rule and using our previous result that $\lim_{c \to 0^-} \frac{\partial A}{\partial c} \to 0$,

[9] This is equivalent to showing that $\lim_{\Delta c \to 0} \frac{A(0+\Delta c) - A(0)}{\Delta c}$ exists (see, for example, Chiang (1984, pp. 149-153)).

one can show

$$\lim_{c \to 0^-} \frac{A(c) - A(0)}{c - 0} = \lim_{c \to 0^-} \frac{\partial A(c)/\partial c}{1} = 0.$$

As the left-side limit equals the right-side limit, therefore $\lim_{c \to 0} \frac{A(c) - A(0)}{c - 0}$ exists and $A(c)$ is differentiable at $c = 0$. Q.E.D.

APPENDIX 5C

Figure 5.1

The minimum risk boundary of $X\hat{\beta}$ combining the unrestricted and inequality restricted estimators.

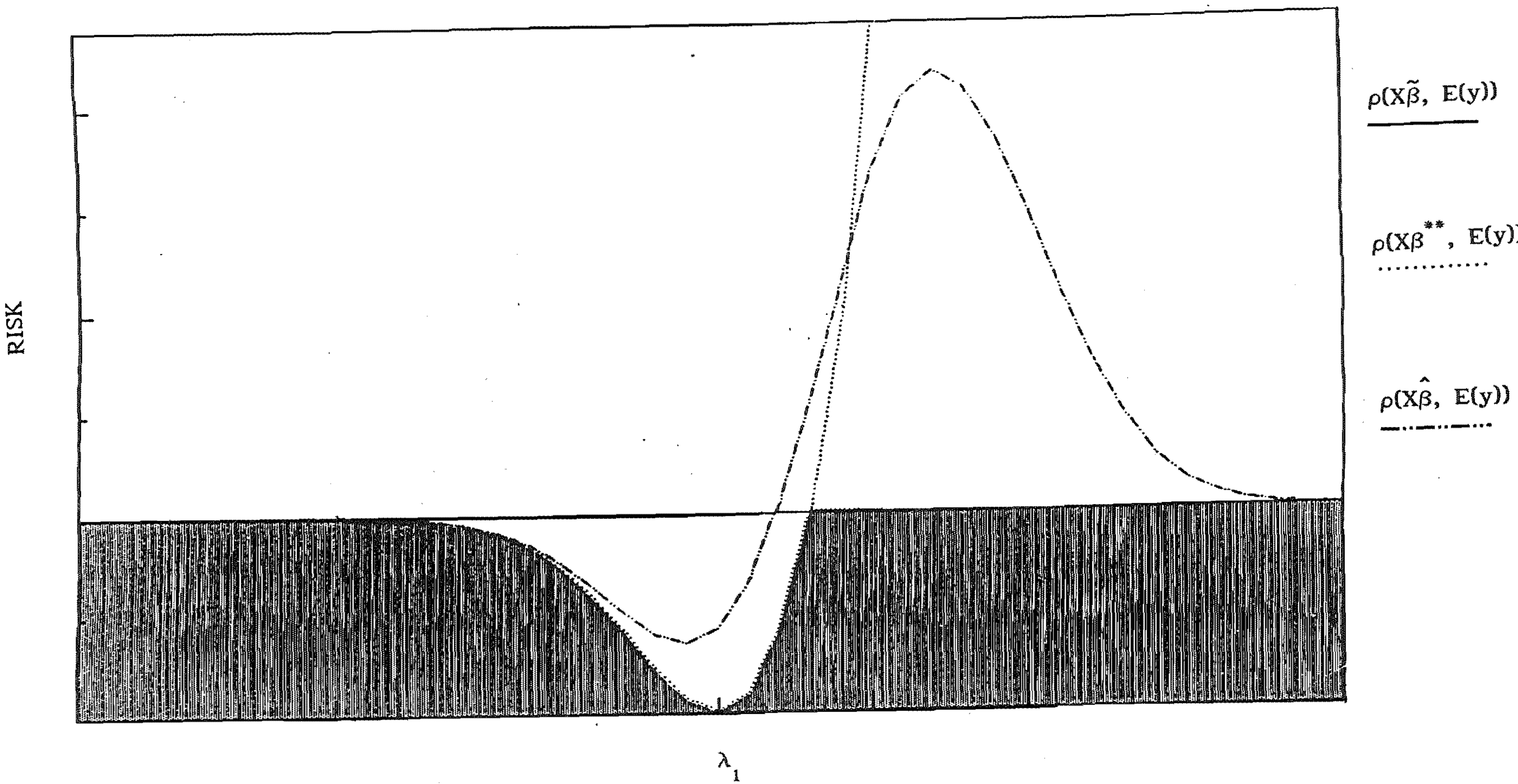


Figure 5.2

$\rho(X\hat{\beta}, E(y))$ for α = 5% and α = 25% (n = 30, k =5 and λ_2 = 0)

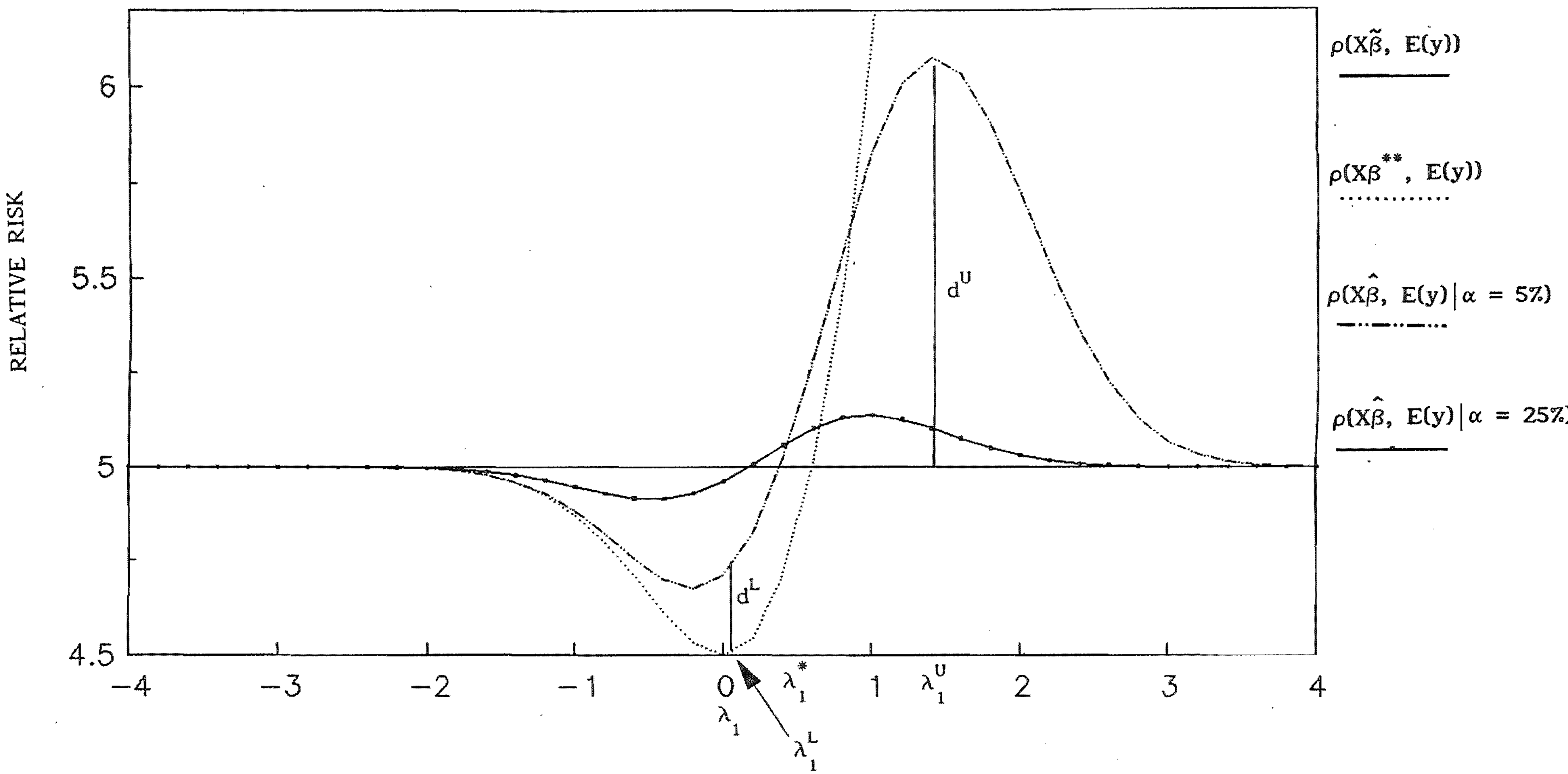


Figure 5.3

$\rho(X\hat{\beta}, E(y))$ for n = 30, k = 5 and $\lambda_2 = 0$, with $c = c^{MX}$ and c = -1.708 (α = 5%)

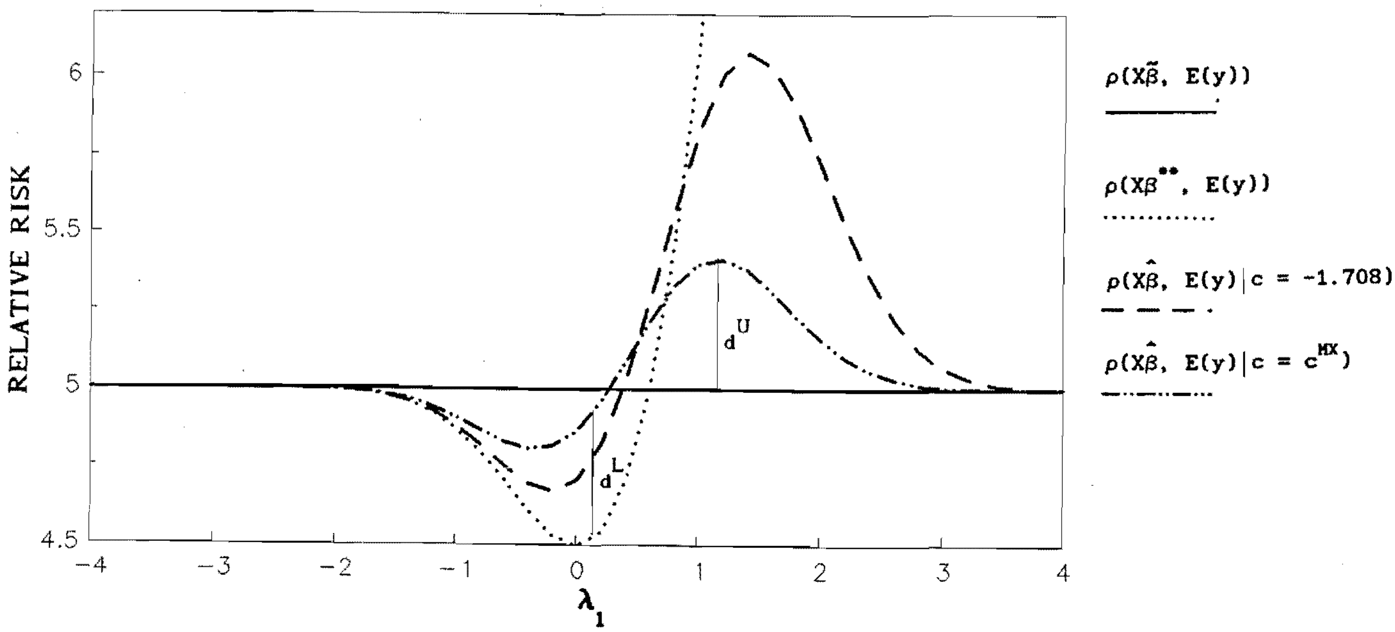


Figure 5.4

$\rho(X\hat{\beta}, E(y))$ for n = 70, k = 10 and $\lambda_2 = 0$, with $c = c^{MX}$ and c = -1.671 (α = 5%)

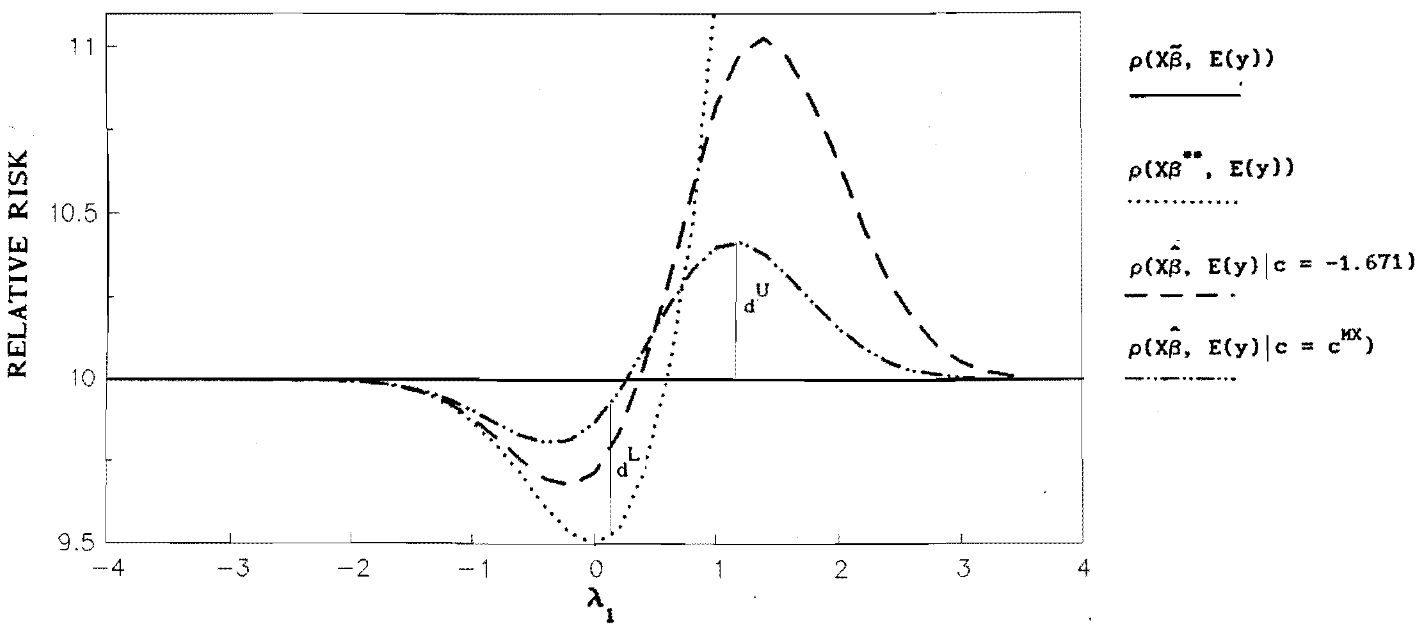


Figure 5.5

$\rho(X\hat{\beta}, E(y))$ for n = 30, k = 5 and λ_2 = 2, with c = c^{MX} and c = -1.127 (α = 5%)

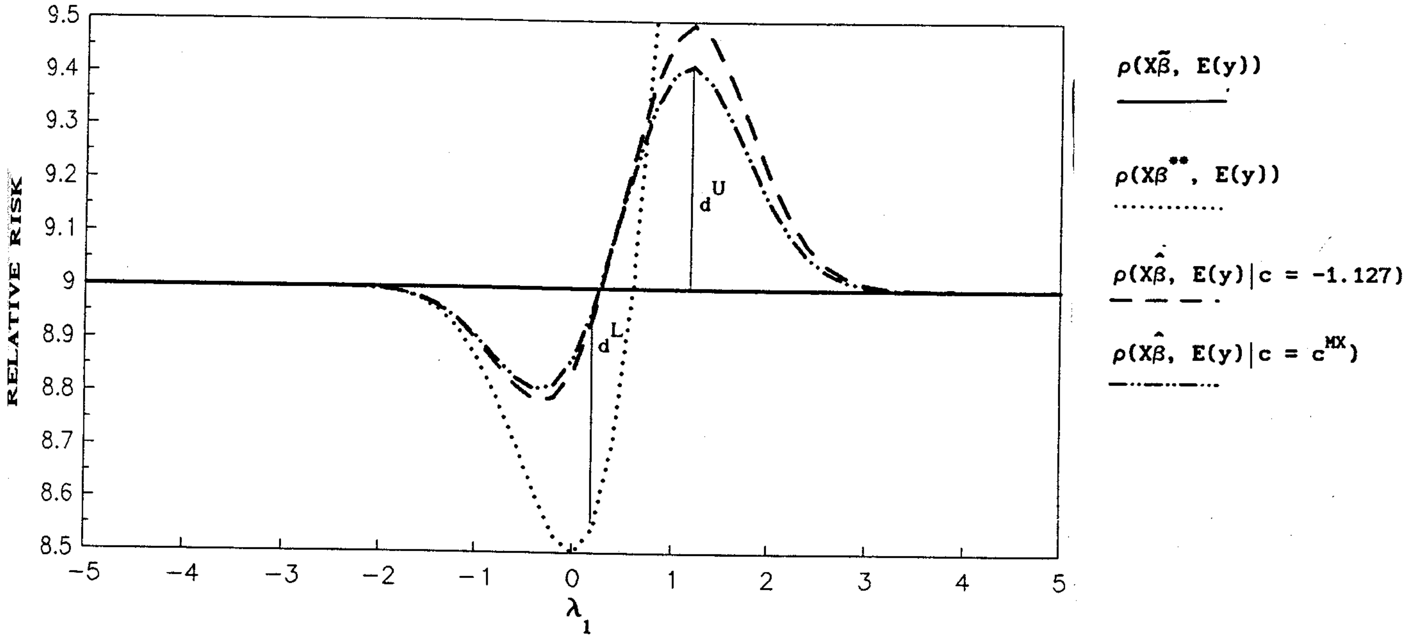


Figure 5.6

$\rho(X\hat{\beta}, E(y))$ for n = 30, k = 5 and λ_2 = 10, with c = c^{MX} and c = -1.127 (α = 5%)

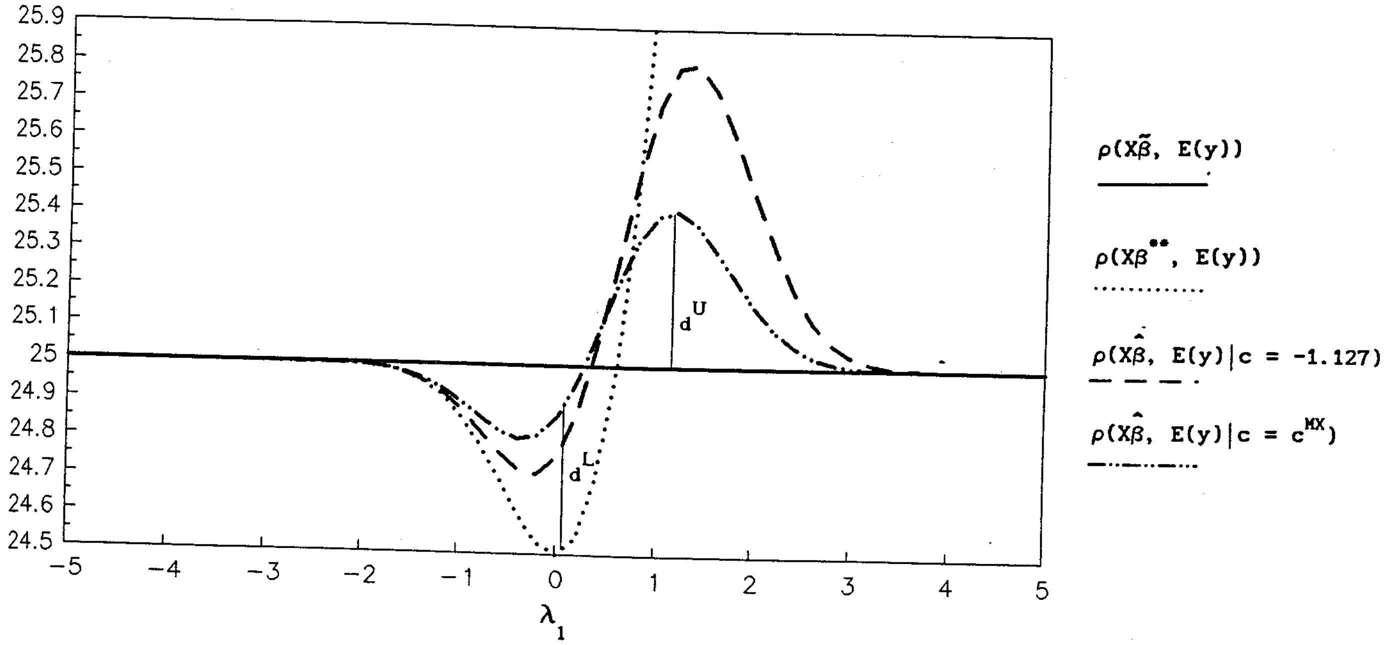


Figure 5.7

$\rho(X\hat{\beta}, E(y))$ for $n = 30$, $k = 5$ and $\lambda_2 = 25$, with $c = c^{MX}$ and $c = -1.127$ ($\alpha = 5\%$)

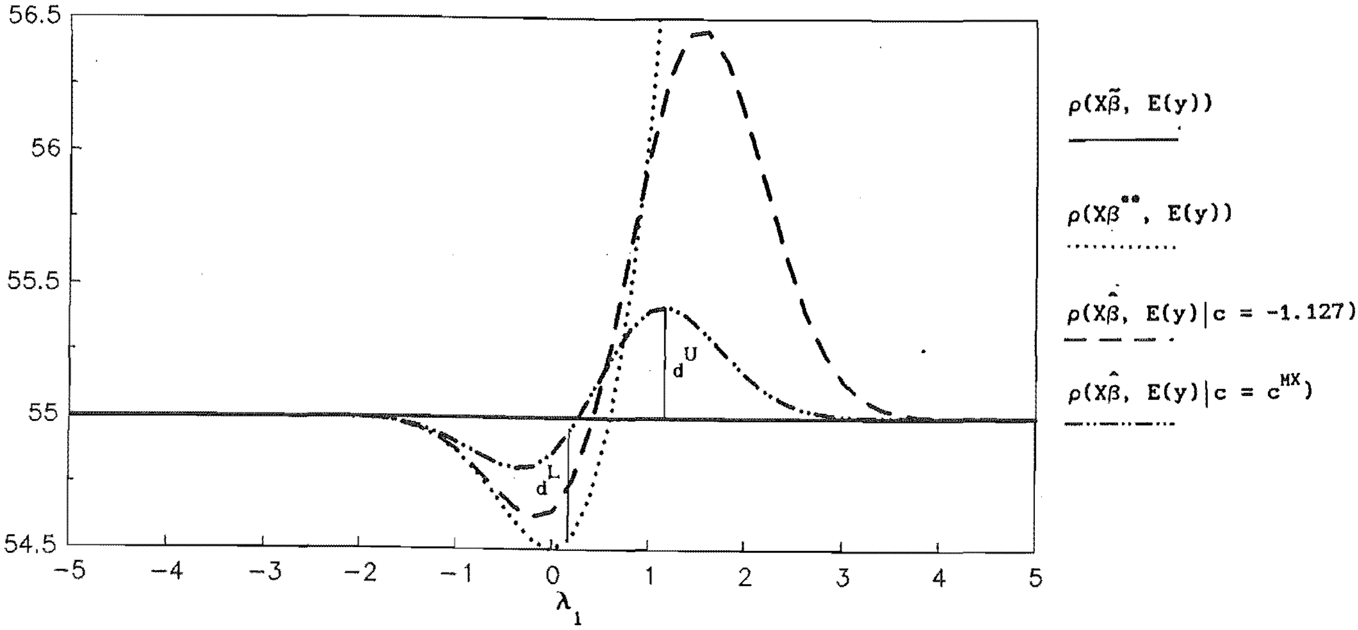


Figure 5.8

$\rho(X\hat{\beta}, E(y))$ for $n = 30$, $k = 5$ and $\lambda_2 = 50$, with $c = c^{MX}$ and $c = -1.127$ ($\alpha = 5\%$)

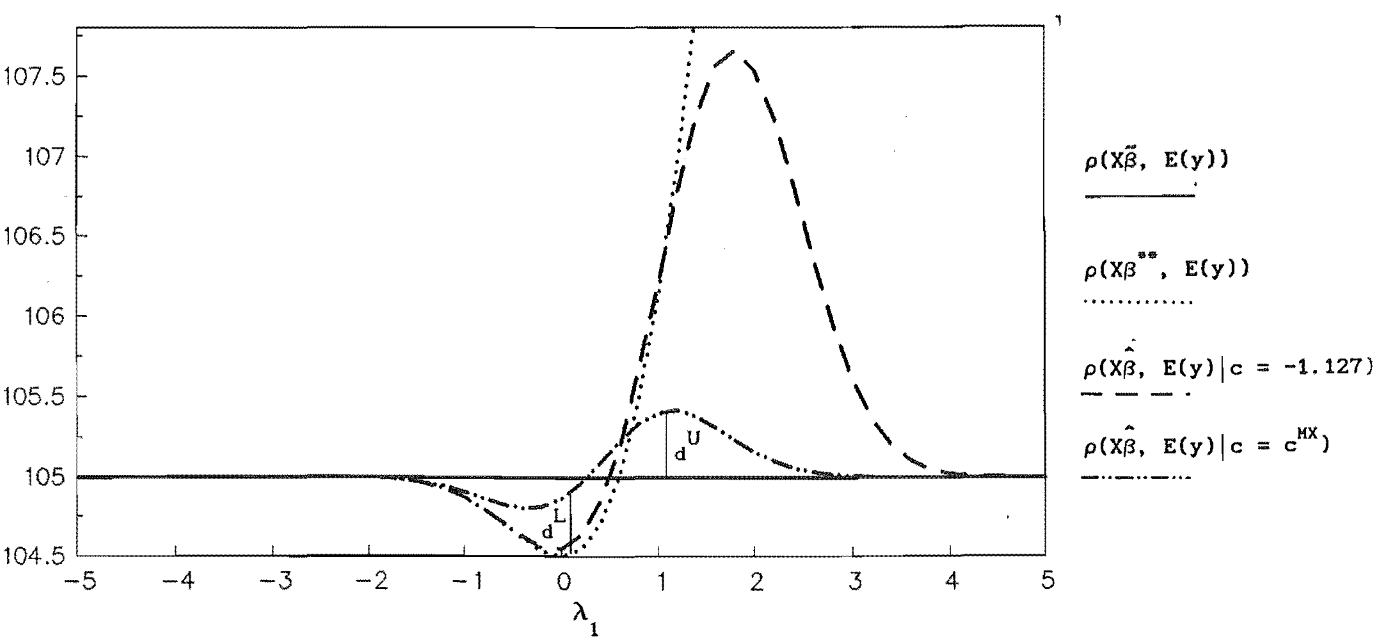


Figure 5.9

$\rho(X\hat{\beta}, E(y))$ for α = 10% and α = 25% (n = 30, k =5 and λ_2 = 0)

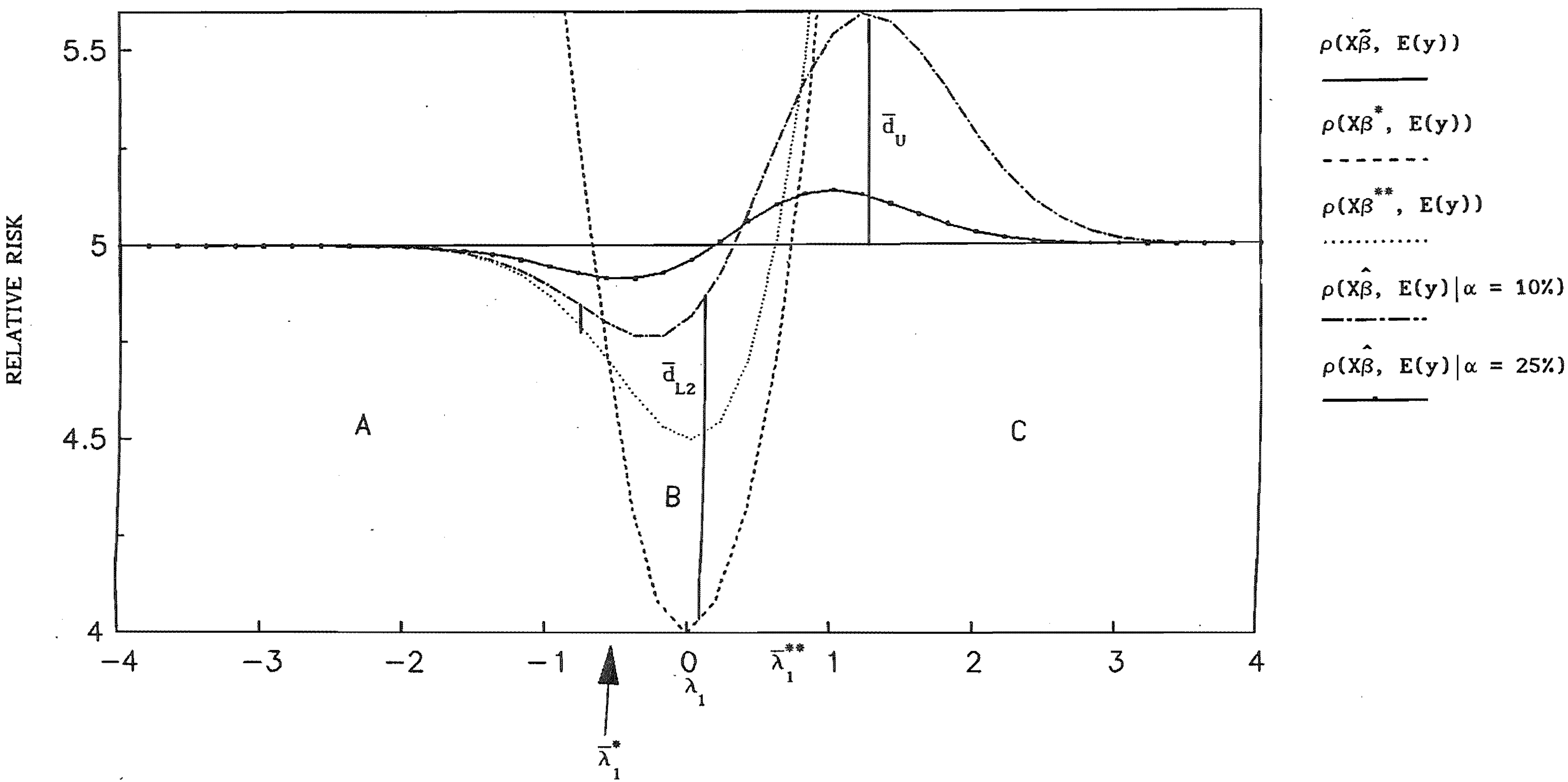


Figure 5.10

$\rho(X\hat{\beta}, E(y))$ for n = 40, k = 5 and λ_2 = 0, with c = c^{MX} and c = c^{MX*}

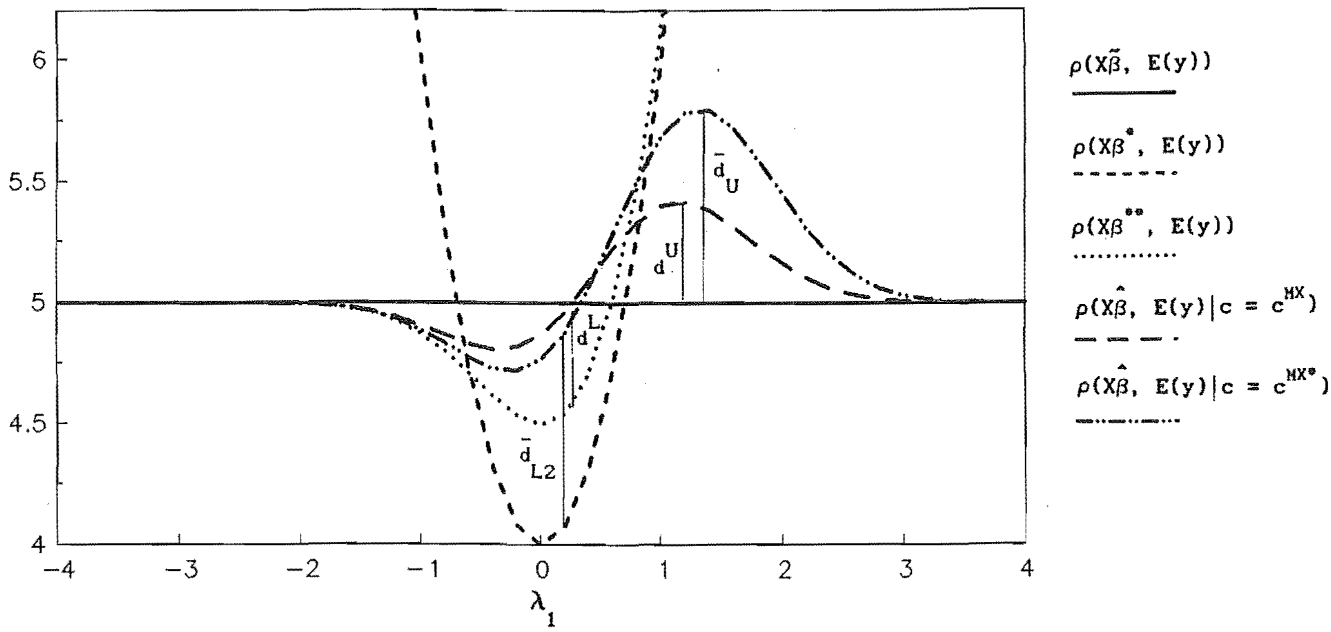


Figure 5.11

$\rho(X\hat{\beta}, E(y))$ for n = 60, k = 10 and λ_2 = 10, with c = c^{MX} and c = c^{MX*}

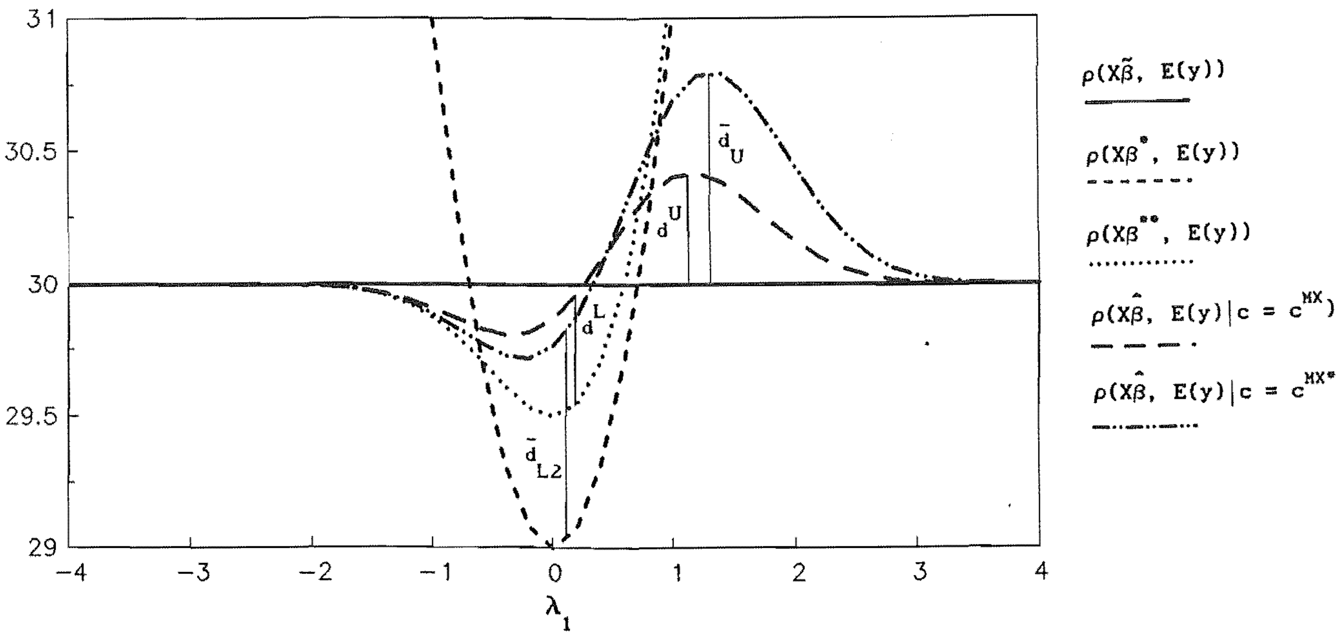


Figure 5.12

Average relative risk of $X\hat{\beta}$ as a function of c for n = 30, k = 5 and λ_2 = 0

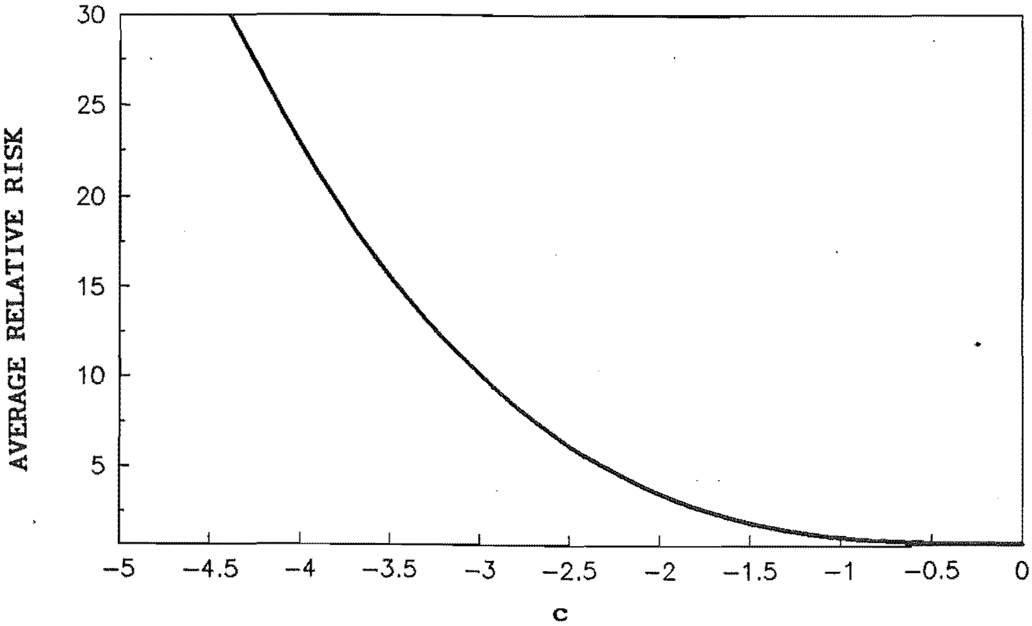
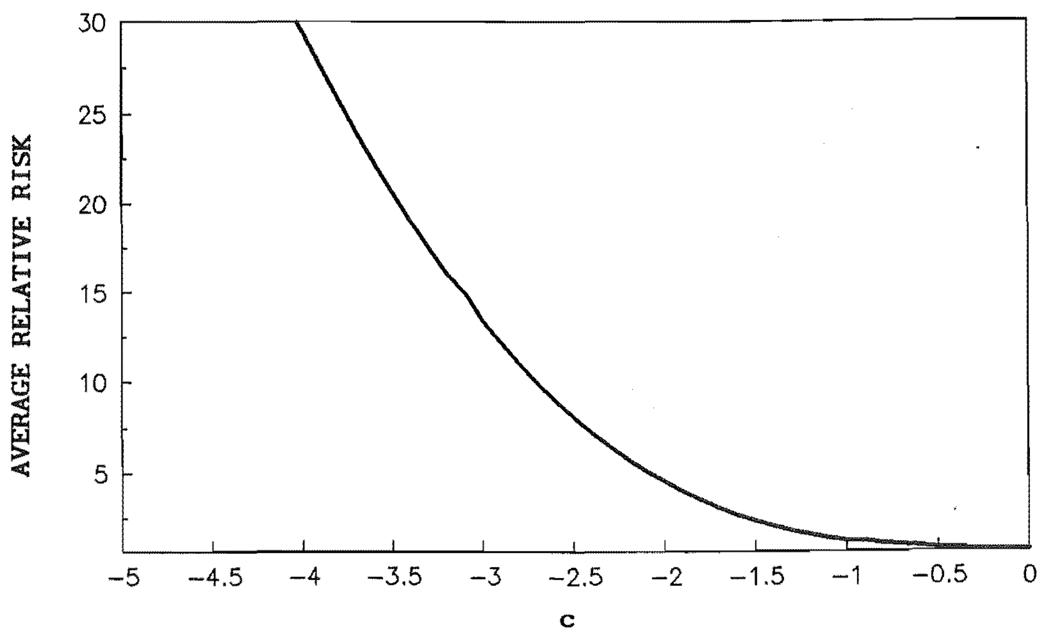


Figure 5.13

Average relative risk of $X\hat{\beta}$ as a function of c for n = 20, k = 2 and λ_2 = 2



CHAPTER SIX

THE SAMPLING PERFORMANCE OF THE INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS FOR THE SCALE PARAMETER IN THE STANDARD LINEAR MODEL

6.1 INTRODUCTION

Traditionally much of the pre-test and shrinkage estimation literature has focussed on estimators of the coefficient or prediction vectors. By way of comparison, the estimation of the scale parameter has received much less attention. As argued by Giles and Giles (1993), this is not surprising as σ^2 is often regarded as a nuisance parameter when interest centers on β. However, in practical terms, the estimation of σ^2 is necessary if the researcher is interested in an analysis of the precision of estimators of β, or if hypothesis tests are to be carried out.

In the context of the standard linear model, Clarke *et al.* (1987a, b) examined the estimation of σ^2 after a pre-test of exact linear equality restrictions on the regression coefficient vector. In this chapter we extend their analysis to the case where the restriction on the regression coefficients is in the form of the single linear inequality hypothesis considered in Chapter 4. Within this framework, we derive and numerically evaluate the risk functions of several inequality restricted and pre-test estimators of σ^2. These estimators are associated with the least squares (LS), maximum likelihood (ML) and the minimum mean squared error (MM) component estimators of the scale parameter.

The rest of this chapter is presented in the following way : In Section 6.2, we detail the derivation of the exact finite sample risk of a general family of inequality restricted estimators of σ^2 under quadratic loss. These risk functions are numerically evaluated for the three component estimators previously mentioned above. In Section 6.3, we focus on the sampling performance of the corresponding inequality pre-test estimators. Finally, some concluding remarks are given in Section 6.4.

6.2 THE RISK PROPOERTIES OF THE INEQUALITY RESTRICTED ESTIMATOR OF σ^2

Consider the standard linear regression model

$$y = X\beta + \varepsilon \quad , \qquad \varepsilon \sim N(0, \sigma^2 I) \tag{6.1}$$

where ε is an $n \times 1$ random vector, y and X are an $n \times 1$ vector and $n \times k$ matrix of observations respectively, with X non-stochastic and of rank k, and β is a $k \times 1$ vector of unknown coefficients.

In addition to the sample information, there exists uncertain prior information regarding β, described by

$$C'\beta \geq r \quad , \tag{6.2}$$

where C is a $k \times 1$ vector of known elements and r is a known scalar.

For purposes of analytical convenience, we follow the approach of Judge and Yancey (1981, 1986) and transform (6.1) and (6.2) into

$$y = H\theta + \varepsilon, \tag{6.3}$$

and

$$\theta_1 \geq r_0 \tag{6.4}$$

respectively, where H, θ_1 and r_0 are defined as in Chapter 3. Furthermore, $H'H = I$.

The unrestricted estimator (UE) of σ^2 is $\tilde{\sigma}^2 = \tilde{e}'\tilde{e}/(n+\delta)$ and the equality restricted estimator (ERE) of σ^2 is $\sigma^{*2} = e^{*\prime}e^*/(n+\gamma)$, where $\tilde{e}$ and e^* are the vectors of residuals corresponding to the unrestricted estimator $\tilde{\theta}$ and the equality restricted estimator θ^* of θ respectively. The maximum likelihood estimators of σ^2 correspond to $\delta = \gamma = 0$. The least squares estimators of σ^2 correspond to $\delta = -k$ and $\gamma = -k+1$, while the minimum mean squared error estimators correspond to $\delta = -k+2$ and $\gamma = -k+3$. Throughout this chapter, the subsripts ML, LS and MM are used to distinguish the estimators that correspond to these components.

Following the convention adopted in the literature, we define the relative risk of an estimator $\bar{\sigma}^2$ of σ^2 as $E\left[(\bar{\sigma}^2 - \sigma^2)^2\right]/\sigma^4$. From Clarke *et al.* (1987b), the relative risk functions of $\tilde{\sigma}^2$ and σ^{*2} are given by

$$\rho(\tilde{\sigma}^2, \sigma^2) = (2v + (k+\delta)^2)/(n+\delta)^2 \tag{6.5}$$

and

$$\rho(\sigma^{*2}, \sigma^2) = \left[2(1+v+2\tau^2/\sigma^2) + (1-k-\gamma+\tau^2/\sigma^2)^2\right]/(n+\gamma)^2 \tag{6.6}$$

respectively, where $v = n - k$ and $\tau = r_0 - \theta_1$ is the surplus variable associated with constraint (6.4).

If the unrestricted estimator of θ satisfies (6.4), then the restriction is believed to be non-binding and the scale parameter is estimated by the unrestricted estimator $\tilde{\sigma}^2$, otherwise the restriction is treated as binding and the scale parameter is estimated by the restricted estimator σ^{*2}. This procedure gives rise to the following inequality restricted estimator (IRE)

$$\sigma^{**2} = \begin{cases} \sigma^{*2} & \text{if } \tilde{\theta}_1 < r_0 \\ \\ \tilde{\sigma}^2 & \text{if } \tilde{\theta}_1 \geq r_0 \end{cases} = I_{(-\infty,r_0)}(\tilde{\theta}_1)\sigma^{*2} + I_{[r_0,\infty)}(\tilde{\theta}_1)\tilde{\sigma}^2 \quad . \tag{6.7}$$

Recognising that $I_{[r_0,\infty)}(.) = 1 - I_{(-\infty,r_0)}(.)$, (6.7) can be rewritten as

$$\sigma^{**2} = \tilde{\sigma}^2 + I_{(-\infty,r_0)}(\tilde{\theta}_1)(\sigma^{*2} - \tilde{\sigma}^2) \quad . \tag{6.8}$$

To simplify (6.8), we recall that $e^* = y - H\theta^*$ and that $\theta^* = \begin{bmatrix} r_0 \\ \tilde{\theta}_{k-1} \end{bmatrix}$.

This implies

$$e^* = y - H\begin{bmatrix} r_0 \\ \tilde{\theta}_{k-1} \end{bmatrix}$$

$$= y - H\tilde{\theta} + H\begin{bmatrix} \tilde{\theta}_1 - r_0 \\ 0 \end{bmatrix}$$

$$= \tilde{e} + H\begin{bmatrix} \tilde{\theta}_1 - r_0 \\ 0 \end{bmatrix} \qquad (6.9)$$

Therefore,

$$e^{*\prime}e^* = \tilde{e}'\tilde{e} + (\tilde{\theta}_1 - r_0)^2 + 2\tilde{e}'H\begin{bmatrix} \tilde{\theta}_1 - r_0 \\ 0 \end{bmatrix} \qquad \text{since } H'H = I$$

$$= \tilde{e}'\tilde{e} + (\tilde{\theta}_1 - r_0)^2 \qquad \text{since } \tilde{e}'H = 0$$

$$= \tilde{e}'\tilde{e} + (\tilde{\theta}_1 - \theta_1 - (r_0 - \theta_1))^2$$

$$= \tilde{e}'\tilde{e} + (\sigma u_1 - \tau)^2 \quad , \qquad (6.10)$$

where $u_1 = (\tilde{\theta}_1 - \theta_1)/\sigma$ is a standard normal variable. Using this result, $\sigma^{*2} - \tilde{\sigma}^2$ may be expressed as

$$\sigma^{*2} - \tilde{\sigma}^2 = \left[\tilde{e}'\tilde{e} + (\sigma u_1 - \tau)^2\right]/(n+\gamma) - \tilde{\sigma}^2$$

$$= \left[(\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right]/(n+\gamma) \quad . \qquad (6.11)$$

Hence,

$$\sigma^{**2} = \tilde{\sigma}^2 + I_{(-\infty, r_0)}(\tilde{\theta}_1)\left[(\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right]/(n+\gamma)$$

$$= \tilde{\sigma}^2 + I_{(-\infty, \tau/\sigma)}(u_1)\left[(\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right]/(n+\gamma) \quad . \qquad (6.12)$$

Using the definition of the relative risk of an estimator as stated earlier, the relative risk of σ^{**2} is

$$\rho(\sigma^{**2}, \sigma^2) = E\left\{\tilde{\sigma}^2 - \sigma^2 + I_{(-\infty, \tau/\sigma)}(u_1)\left[\left((\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right]\right\}^2/\sigma^4$$

$$= \rho(\tilde{\sigma}^2, \sigma^2) + E\left\{I_{(-\infty, \tau/\sigma)}(u_1)\left[\left((\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right]^2\right\}/\sigma^4$$

$$+ 2E\left\{(\tilde{\sigma}^2 - \sigma^2)I_{(-\infty, \tau/\sigma)}(u_1)\left[\left((\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right]\right\}/\sigma^4$$

$$= \rho(\tilde{\sigma}^2, \sigma^2) + E\left\{I_{(-\infty, \tau/\sigma)}(u_1)\left[\left((\sigma u_1 - \tau)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right]\right.$$

$$\times\left[2(\tilde{\sigma}^2-\sigma^2) + \left((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\right)/(n+\gamma)\right]\Big\}/\sigma^4 . \qquad (6.13)$$

In order to evaluate this risk, we need to evaluate $E\left[\tilde{\sigma}^i I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\right]$ for i = 0, 2, 4 and j = 0, 1, 2, 3, 4. Now, u_1 is distributed independently[1] of $\tilde{\sigma}^2$. Hence $E\left[\tilde{\sigma}^i I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\right] = E(\tilde{\sigma}^i)E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\right]$. It is well known that $(n+\delta)\tilde{\sigma}^2/\sigma^2 \sim \chi_v^2$. From the moments of a Chi-Square random variable, $E(\chi_v^2) = v$ and $E(\chi_v^4) = v(v+2)$. Accordingly,

$$E(\tilde{\sigma}^2) = \sigma^2 v/(n+\delta) \qquad (6.14)$$

and

$$E(\tilde{\sigma}^4) = \sigma^4 v(v+2)/(n+\delta)^2. \qquad (6.15)$$

Now, $E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\right]$, j = 0, 1, 2, 3, 4, may be evaluated using the following Corollaries :

<u>Corollary 6.1</u>

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)\right] = \begin{cases} \frac{1}{2}P(\chi_1^2 \ge \tau^2/\sigma^2) & \text{if } \tau \le 0 \\ \\ 1-\frac{1}{2}P(\chi_1^2 \ge \tau^2/\sigma^2) & \text{if } \tau > 0 \end{cases}, \qquad (6.16)$$

<u>Corollary 6.2</u>

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1\right] = -\frac{1}{\sqrt{2\pi}}P(\chi_2^2 \ge \tau^2/\sigma^2) , \qquad (6.17)$$

<u>Corollary 6.3</u>

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)\right] = \begin{cases} \frac{1}{2}P(\chi_3^2 \ge \tau^2/\sigma^2) & \text{if } \tau \le 0 \\ \\ 1-\frac{1}{2}P(\chi_3^2 \ge \tau^2/\sigma^2) & \text{if } \tau > 0 \end{cases}, \qquad (6.18)$$

---

[1] Since $\tilde{\theta}$ and $\tilde{\sigma}^2$ are independently distributed, it follows that u_1 is also distributed independently of $\tilde{\sigma}^2$.

Corollary 6.4

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^3\right] = -\sqrt{\frac{2}{\pi}}\,P(\chi_4^2 \geq \tau^2/\sigma^2) \quad , \tag{6.19}$$

Corollary 6.5

$$E\left[I_{(-\infty,\tau/\sigma)}(u_1)u_1^4\right] = \begin{cases} \frac{3}{2}\,P(\chi_5^2 \geq \tau^2/\sigma^2) & \text{if } \tau \leq 0 \\ 3 - \frac{3}{2}\,P(\chi_5^2 \geq \tau^2/\sigma^2) & \text{if } \tau > 0 \end{cases} \quad . \tag{6.20}$$

Proof : These corollaries follow from Theorem 1 of Judge and Yancey (1986, Chapter 4) by setting j = 0, 1, 2, 3 and 4 respectively. Corollaries 6.1 - 6.3 were previously stated in Judge and Yancey (1986, pp. 74-75).

Making use of these corollaries in conjunction with (6.13), (6.14) and (6.15), and after performing some tedious algebraic manipulations,[2] we can show that the risk of σ^{**2}, when $\tau \leq 0$, and thus the inequality restriction is valid, is

$$\begin{aligned}\rho(\sigma^{**2},\ \sigma^2) = {} & \rho(\tilde{\sigma}^2,\ \sigma^2) + \Big\{(\tau/\sigma)^2(n+\delta)^2\Big[(\tau/\sigma)^2 - 2(n+\gamma)\Big] + 2v(n+\delta)\Big[(\tau/\sigma)^2(n+\delta) \\ & + (\gamma-\delta)(n+\gamma)\Big] - v(v+2)(\gamma-\delta)(2n + \gamma + \delta)\Big\}P_1/\Big[2((n+\delta)(n+\gamma))^2\Big] + \\ & \Big[4(\tau/\sigma)^3 + 4(\tau/\sigma)(v - (n+\gamma))\Big]P_2/\Big[\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + \Big[v+3(\tau/\sigma)^2 - \\ & (n+\gamma)\Big]P_3/(n+\gamma)^2 + 8(\tau/\sigma)P_4/\Big[\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + 3P_5/(2(n+\gamma)^2) \quad .\end{aligned} \tag{6.21}$$

Alternatively, when $\tau > 0$, and thus the direction of the inequality constraint

---

[2] Full details are available upon request.

is incorrect, the risk of σ^{**2} is, by applying (6.14), (6.15) and Corollaries 6.1 - 6.5 again,

$$\rho(\sigma^{**2}, \sigma^2) = \rho(\tilde{\sigma}^2, \sigma^2) + \Bigg\{(\tau/\sigma)^2(n+\delta)^2\Big[(\tau/\sigma)^2 - 2(n+\gamma)\Big] + 2v(n+\delta)\Big[(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)(n+\gamma)\Big] - v(v+2)(\gamma-\delta)(2n + \gamma + \delta)\Bigg\}\Big[1 - P_1/2\Big]/ \Big[((n+\delta)(n+\gamma))^2\Big] + \Big[4(\tau/\sigma)^3 + 4(\tau/\sigma)(v - (n+\gamma))\Big]P_2/\Big[\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + \Big[v+3(\tau/\sigma)^2 - (n+\gamma)\Big](2 - P_3)/(n+\gamma)^2 + 8(\tau/\sigma)P_4/\Big[\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + (6 - 3P_5)/(2(n+\gamma)^2) \tag{6.22}$$

where $P_I = P(\chi^2_I \geq \tau^2/\sigma^2)$, $I = 1,..,5$.

From the Convergence Theorem of Judge and Yancey (1986, p. 77), $\tau P_I \rightarrow 0$ as $|\tau| \rightarrow \infty$ for all I. Hence $\rho(\sigma^{**2}, \sigma^2)$ approaches $\rho(\tilde{\sigma}^2, \sigma^2)$ as $\tau \rightarrow -\infty$, and is asymptotic to $\rho(\tilde{\sigma}^2, \sigma^2) + \Big\{(\tau/\sigma)^2(n+\delta)^2\Big[(\tau/\sigma)^2 - 2(n+\gamma)\Big] + 2v(n+\delta)\Big[(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)(n+\gamma)\Big] - v(v+2)(\gamma-\delta)(2n + \gamma + \delta)\Big\}/\Big[(n+\delta)(n+\gamma)\Big]^2 + \Big\{2\Big[v + 3(\tau/\sigma)^2 - (n+\gamma)\Big] + 3\Big\}/(n+\gamma)^2$ as $\tau \rightarrow \infty$. We prove in Appendix 6A that this is in fact the risk of the ERE. Intuitively, this result arises because when τ is infinitely large, the likelihood of the UE violating the inequality constraint is high. Consequently, the ERE is chosen as the estimator of the model a larger proportion of the time. The converse is true when τ is infinitely small. This is consistent with the corresponding result when estimating E(y) or β. Given the complexity of these risk functions, it is difficult to analyse them further without undertaking some numerical evaluations.

Numerical calculations of (6.21) and (6.22) have been carried out when n = 20, 30, 40, 50, 80, k = 2, 5, 10, 15,...n-5 and $\tau/\sigma \in [-10, 10]$. They were performed on a VAX 6340 computer using double precision FORTRAN code which incorporates the subroutine GAMMQ given in Press *et al.* (1984) to evaluate P_I

for I = 1, 2,..,5. Some representative diagrams illustrating these risks are given in Appendix 6B. The risks of the unrestricted and equality restricted estimators are also shown in these diagrams as reference functions for comparison. As in the case of estimating E(y) and β, we show numerically that when estimating σ^2, the risk of the IRE is always smaller than that of the UE when $\tau \le 0$. However, unlike the case of estimating the coefficent or prediction vector, the biggest risk improvement of using σ^{**2} over $\tilde{\sigma}^2$ does *not* necessarily occur at $\tau = 0$. The minimum of $\rho(\sigma^{**2}, \sigma^2)$ tends to occur when $\tau < 0$ if the component estimator is LS or MM, and in the positive horizon of τ if the component is ML. Over a large portion of the parameter space, $\tilde{\sigma}^2_{ML}$ is dominated by σ^{**2}_{ML}. By comparison, the region over which σ^{**2}_{MM} and σ^{**2}_{LS} dominate the corresponding unrestricted estimators of their respective families is typically smaller. Regardless of the choice of the component estimator, the risk of the IRE intersects that of the UE when $\tau > 0$ and increases without bound thereafter. Again, this accords with the corresponding result that is observed when estimating E(y) or β.

6.3 THE INEQUALITY PRE-TEST ESTIMATOR OF σ^2

As in Chapter 3, the inequality hypothesis (6.4) is tested typically using the statistic

$$t = \sqrt{v}(\tilde{\theta}_1 - r_0)/(\tilde{\sigma}\sqrt{(n+\delta)}) \qquad (6.23)$$

When $\theta = r$, t has a Student's *t* distribution with v degrees of freedom. We reject the null if $t \le c$ and use the unrestricted estimator, where c is the size α critical value obtained from the Student's *t* table. We do not reject the null if $t > c$ and we then use the inequality restricted estimator. Accordingly, this mechanism gives rise to the inequality pre-test estimator (IPTE)

$$\hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2 & \text{if } t < c \\ \sigma^{**2} & \text{if } t \geq c \end{cases} = I_{(-\infty,c)}(t)\tilde{\sigma}^2 + I_{[c,\infty)}(t)\sigma^{**2} \quad . \tag{6.24}$$

Again, as in Chapter 4, the case of $c \geq 0$ needs no discussion.[3] In the analysis that follows, we assume that $c < 0$.

Using the usual properties of indictor functions and the results from (6.8), (6.24) can be simplified to

$$\hat{\sigma}^2 = \sigma^{**2} - I_{(-\infty,c)}(t)\Big[I_{(-\infty,\tau/\sigma)}(u_1)(\sigma^{*2} - \tilde{\sigma}^2)\Big] \ . \tag{6.25}$$

From (6.11), $\sigma^{*2} - \tilde{\sigma}^2 = \Big((\sigma u_1-\tau)^2 - \tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)$. Accordingly,

$$\hat{\sigma}^2 = \sigma^{**2} - I_{(-\infty,c)}(t)\Big[I_{(-\infty,\tau/\sigma)}(u_1)\Big((\sigma u_1-\tau)^2 - \tilde{\sigma}^2(\gamma-\delta)\Big)\Big]/(n+\gamma) \ . \tag{6.26}$$

Now, $I_{(-\infty,c)}(t)I_{(-\infty,\tau/\sigma)}(u_1) \quad = I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)I_{(-\infty,\tau/\sigma)}(u_1)$

$$= I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1), \qquad \text{since } c' < 0,$$

where $c' = c\sqrt{(n+\delta)/v}$. Accordingly,

$$\hat{\sigma}^2 = \sigma^{**2} - I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\Big[(\sigma u_1-\tau)^2 - \tilde{\sigma}^2(\gamma-\delta)\Big]/(n+\gamma) \ . \tag{6.27}$$

By analogy with (6.13), applying (6.12) and using the fact that $I_{(-\infty,c)}(t)I_{(-\infty,\tau/\sigma)}(u_1) = I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)$, the relative risk of $\hat{\sigma}^2$ may be written as

$$\rho(\hat{\sigma}^2, \sigma^2) = \rho(\sigma^{**2}, \sigma^2) - E\Big\{I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]$$
$$\times\Big[2(\tilde{\sigma}^2-\sigma^2) + \Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \ . \tag{6.28}$$

In order to analyse (6.28), we need to evaluate $E\Big[\tilde{\sigma}^i I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\Big]$, $i = 0, 2, 4$ and $j = 0, 1, 2, 3, 4$. In evaluating the former, we require the

---

[3] See Chapter 4, p. 86 for explanations and details.

following corollaries which follow from Theorem 6.1 given in Appendix 6A at the end of this chapter.

Corollary 6.6

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\right] = \begin{cases} \dfrac{\bar{E}_{1,v}}{2} & \text{if } \tau \le 0 \\ \dfrac{\bar{E}_{1,v}}{2} + \bar{G}_{1,v} & \text{if } \tau > 0 \end{cases} \quad , \qquad (6.29)$$

Corollary 6.7

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1\right] = -\frac{1}{\sqrt{2\pi}}\bar{E}_{2,v} \quad , \qquad (6.30)$$

Corollary 6.8

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^2\right] = \begin{cases} \dfrac{\bar{E}_{3,v}}{2} & \text{if } \tau \le 0 \\ \dfrac{\bar{E}_{3,v}}{2} + \bar{G}_{3,v} & \text{if } \tau > 0 \end{cases} \quad , \qquad (6.31)$$

Corollary 6.9

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^3\right] = -\sqrt{\frac{2}{\pi}}\,\bar{E}_{4,v} \quad , \qquad (6.32)$$

Corollary 6.10

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^4\right] = \begin{cases} \dfrac{3\bar{E}_{5,v}}{2} & \text{if } \tau \le 0 \\ \dfrac{3\bar{E}_{5,v}}{2} + 3\bar{G}_{5,v} & \text{if } \tau > 0 \end{cases} \quad , \qquad (6.33)$$

where

$$\bar{E}_{I,J} = \frac{1}{2^{J/2}\Gamma(\frac{J}{2})}\int_0^\infty P(\chi_I^2 \ge (cq_J/\sqrt{v}+\tau)^2)\,(q_J^2)^{J/2-1}\,e^{-q_J^2/2}\,dq_J^2,$$

$$\bar{G}_{I,J} = \frac{1}{2^{J/2}\Gamma(\frac{J}{2})}\int_0^{v\tau^2/c^2} P(\chi_I^2 < (cq_J/\sqrt{v}+\tau)^2)\ (q_J^2)^{J/2-1}\ e^{-q_J^2/2}\ dq_J^2,$$

$I = 1, 2, 3, 4, 5$ and q_J^2 is a Chi-Square random variable with J degrees of freedom. Note also that $I = j + 1$.

The evaluation of $E\left[\tilde{\sigma}^i I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$ is slightly more complicated as $\tilde{\sigma}^2$ and $I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j$ are clearly not independently distributed. However, as $\tilde{\sigma}$ is non-negative by definition, for each $\tilde{\sigma}^2$, there is only one corresponding $\tilde{\sigma}$. Hence, $I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j|u_1$ can be regarded as a function of $\tilde{\sigma}^2$. Now, recall that $(n+\delta)\tilde{\sigma}^2/\sigma^2 \sim \chi_v^2$. According to Theorem 2 of Judge and Bock (1978, p. 322), for any real measurable function Φ,

$$E(\chi_v^2\Phi(\chi_v^2)) = vE(\Phi(\chi_{v+2}^2)) \quad . \tag{6.34}$$

Applying this theorem,

$$E\left[\tilde{\sigma}^2 I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right] = \frac{\sigma^2 v}{n+\delta}\, E\left[I_{(-\infty,cq_{v+2}/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right] \tag{6.35}$$

and

$$\begin{aligned} E\left[\tilde{\sigma}^4 I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right] &= \frac{\sigma^4}{(n+\delta)^2}\left\{vE\left[I_{(-\infty,cq_{v+2}/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j q_{v+2}^2\right]\right\} \\ &= \frac{\sigma^4 v(v+2)}{(n+\delta)^2}\, E\left[I_{(-\infty,cq_{v+4}/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right]. \end{aligned} \tag{6.36}$$

Making use of Corollaries 6.6 - 6.10, (6.35) and (6.36), by collecting terms and after performing some tedious manipulations, the relative risk of $\hat{\sigma}^2$, when $\tau \le 0$, may be expressed as

$$\begin{aligned} \rho(\hat{\sigma}^2, \sigma^2) = \ & \rho(\sigma^{**2}, \sigma^2) - 3\bar{E}_{5,v}/2(n+\gamma)^2 - 8(\tau/\sigma)\bar{E}_{4,v}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 + \Big[n+\gamma \\ & -3(\tau/\sigma)^2\Big]\bar{E}_{3,v}/(n+\gamma)^2 + \Big[4(\tau/\sigma)(n+\gamma)-4(\tau/\sigma)^3\Big]\bar{E}_{2,v}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 \\ & + \Big[2(\tau/\sigma)^2(n+\gamma)-(\tau/\sigma)^4\Big]\bar{E}_{1,v}/2(n+\gamma)^2 - v\bar{E}_{3,v+2}/(n+\gamma)^2 - 4v(\tau/\sigma) \\ & \times \bar{E}_{2,v+2}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 - \Big[v(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)v(n+\gamma)\Big]\bar{E}_{1,v+2} \end{aligned}$$

$$/(n+\delta)(n+\gamma)^2 + (\gamma-\delta)v(v+2)(2n+\gamma+\delta)\bar{E}_{1,v+4}/(2(n+\delta)^2(n+\gamma)^2). \qquad (6.37)$$

Alternatively, if $\tau > 0$, the relative risk of $\hat{\sigma}^2$, again by using Corollaries 6.6 - 6.10, (6.35) and (6.36), can be written as

$$\begin{aligned}\rho(\hat{\sigma}^2, \sigma^2) = {} & \rho(\sigma^{**2}, \sigma^2) - 3(\bar{E}_{5,v}/2 + \bar{G}_{5,v})/(n+\gamma)^2 - 8(\tau/\sigma)\bar{E}_{4,v}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 \\ & + \left[n+\gamma -3(\tau/\sigma)^2\right](\bar{E}_{3,v} + 2\bar{G}_{3,v})/(n+\gamma)^2 + \left[4(\tau/\sigma)(n+\gamma)-4(\tau/\sigma)^3\right] \\ & \times \bar{E}_{2,v}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 + \left[2(\tau/\sigma)^2(n+\gamma)-(\tau/\sigma)^4\right](\bar{E}_{1,v}/2 + \bar{G}_{1,v})/(n+\gamma)^2 \\ & - v(\bar{E}_{3,v+2} + 2\bar{G}_{3,v+2})/(n+\gamma)^2 - 4v(\tau/\sigma)\bar{E}_{2,v+2}/\sqrt{2}\Gamma(\tfrac{1}{2})(n+\gamma)^2 - \\ & \left[v(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)v(n+\gamma)\right](\bar{E}_{1,v+2} + 2\bar{G}_{1,v+2})/(n+\delta)(n+\gamma)^2 + \\ & (\gamma-\delta)v(v+2)(2n+\gamma+\delta)(\bar{E}_{1,v+4} + 2\bar{G}_{1,v+4})/(2(n+\delta)^2(n+\gamma)^2). \qquad (6.38)\end{aligned}$$

Note that as $|\tau| \to \infty$, $\bar{E}_{IL} \to 0$ and $\bar{G}_{IL} \to 1$. Using this result, we can show that $\rho(\hat{\sigma}^2, \sigma^2)$ approaches $\rho(\tilde{\sigma}^2, \sigma^2)$ as[4] $|\tau| \to \infty$. Intuitively, this is because when τ is sufficiently large, the likelihood of rejecting the null is high and the UE is chosen more frequently. Alternatively, when τ is sufficently small, the chances of accepting the validity of the inequality restriction is high, and so are the chances of the UE not violating the constraint. Again, this increases the proportion of times that the UE is chosen. Furthermore, when $c \to 0^-$, $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\tilde{\sigma}^2, \sigma^2)$ as $\bar{E}_{IL} \to P_I$ and $\bar{G}_{IL} \to 1 - P_I$; when $c \to -\infty$, $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\sigma^{**2}, \sigma^2)$ as $\bar{E}_{IL}$ and $\bar{G}_{IL}$ both approach zero. These results are consistent with those reported in the literature for the case when one is estimating β.

Numerical evaluations of the risk of $\hat{\sigma}^2$ were undertaken for the same parameter values of n and k as in the previous section, and for $\alpha = 0.01$, 0.05,

---

[4] See Appendix 6A for details.

0.10, 0.25 and 0.40. The subroutine D01AJF from the NAG (1991) Subroutine library, and the subroutine GAMMQ from Press *et al.* (1986) were used to evaluate the integrals $\bar{E}_{IJ}$ and $\bar{G}_{IJ}$. These risks are also depicted in the diagrams of Appendix 6B.

Our numerical results show that regardless of the choice of component estimators, when $\tau \le 0$ and thus the direction of the inequality constraint is correct, the IPTE is superior to the UE, but it is dominated by the IRE. When the IPTE is based on the LS or MM component estimators, there always exists a class of IPTEs that *strictly* dominate the UE *irrespective* of the model's degrees of freedom. Over certain regions in the parameter space, this class of IPTEs also dominate both the UE and IRE simultaneously. Although this class of IPTEs is dominated by the IRE when $\tau \le 0$, with an appropriate choice of test size, the degree of dominance is typically very minor. Among the class of IPTEs that strictly dominate the UE, we find numerically that the estimator with $c = -\sqrt{v/(v+2)}$ (for MM) or $c = -1$ (for LS) always has the smallest risk in the region where $\hat{\sigma}^2$ dominates both $\tilde{\sigma}^2$ and σ^{**2}. We also find that the IPTE with $c > -1$ (for LS) or $c > -\sqrt{v/(v+2)}$ (for MM) are inadmissable as they are dominated by the IPTE with $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM) everywhere.

However, there are some regions where the IPTE with $c < -1$ (for LS) or $c < -\sqrt{v/(v+2)}$ (for MM) has risk smaller than that of the IPTE with $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM). These results concur with those reported by Ohtani (1991a)[5] where one's concern is focused on the estimation of σ^2 after the one sided pre-test of the mean in a normal population,[6] and the resulting pre-test

---

[5] The estimators that Ohtani considered were only those associated with the minimum mean squared error component estimators. He did not consider the estimators of the least squares family.

[6] In constrast to our numerical results, Ohtani obtained his results using analytical methods.

estimator is a choice between the unrestricted and the equality restricted estimators of σ^2. We also show analytically that the risk of $\hat{\sigma}^2$ reaches a stationary point at $c = -1$ for the case of LS and at $c = -\sqrt{v/(v+2)}$ for the case[7] of MM. Again, this result is consistent with those found by Giles (1990, 1991a) for the case in which the linear restriction is in the form of a strict equality.[8]

When estimating σ^2 based on the method of maximum likelihood, our results show that for small k (say, less than 5), the risk of $\hat{\sigma}^2_{ML}$ is always larger than the minimum of $\rho(\tilde{\sigma}^2_{ML}, \sigma^2)$ and $\rho(\sigma^{**2}_{ML}, \sigma^2)$. On the other hand, when k is relatively large, $\hat{\sigma}^2_{ML}$ can strictly dominate $\tilde{\sigma}^2_{ML}$. This feature constrasts with the results found when the linear restriction holds as a strict equality (see Clarke *et al* (1987a)). However, $\hat{\sigma}^2_{ML}$ does not dominate both $\tilde{\sigma}^2_{ML}$ and σ^{**2}_{ML} simultaneously. That is, in the case where $\hat{\sigma}^2_{ML}$ approaches $\tilde{\sigma}^2_{ML}$ from below, $\hat{\sigma}^2_{ML}$ always converges to $\tilde{\sigma}^2_{ML}$ to the left of the intersection between $\tilde{\sigma}^2_{ML}$ and σ^{**2}_{ML}. Compared with the corresponding LS or MM component estimators, σ^{**2}_{ML} generally dominates $\hat{\sigma}^2_{ML}$ over a wider range of the parameter space.

6.4 CONCLUSIONS

In this chapter, we have derived and numerically evaluated the risk functions of the inequality restricted and pre-test estimators of σ^2 in the standard linear model, and focussed our attention on the least squares, maximum likelihood and minimum mean squared error component estimators.

---

[7] This result is proved in Appendix 7A of the next chapter, as it also applies to the case where variables are omitted from the model.

[8] See also Ohtani (1991a).

Regardless of the choice of the component estimators, our results show that there is always a region of τ over which it is better to impose than to ignore the prior information. However, as τ gets larger, the risk of σ^{**2} increases monotonically with τ and is dominated by $\tilde{\sigma}^2$ over a wide region of the parameter space. On the other hand, there exist certain sizes for the pre-test such that the risk of the IPTE based on the LS or MM components can be uniformly smaller than that of the UE. Although this class of IPTE is dominated by the IRE over the region in which the constraint is sufficiently true, the degree of dominance is typically very slight. By comparison, when estimating using the maximum likelihood components, the IPTE is dominated by the IRE estimator over a relatively wider range of the parameter space. With an appropriate choice of test size, $\hat{\sigma}^2_{ML}$ uniformly dominates $\tilde{\sigma}^2_{ML}$ when k is sufficently large.

These findings clearly indicate that when estimating the scale parameter, ignoring the restriction is not recommended. Provided that the model is properly specified, pre-testing is generally the preferred strategy. The question of choosing an optimal size for the pre-test remains and will be addressed in Chapter 8. In the next chapter, we will examine the robustness of the results reported here to mis-specification of the regressor matrix.

APPENDIX 6A

Proof of Corollaries 6.3 to 6.7

As a central Chi-square random variable is just a non-central Chi-square variable with the non-centrality parameter equal to zero, Theorem 4.2 given in Chapter 4 can be applied to evaluate $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right] = E\left[I_{(-\infty,cq_J/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right]$. Now, if $c < 0$, $u_1 \sim N(0,1)$ and $q_J^2 \sim \chi^2_J$, then by applying Theorem 4.2 of Chapter 4, we obtain the following special case :

THEOREM 6.1

If $\tau \le 0$,

$$E\left[I_{(-\infty,cq_J/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right] = \frac{1}{2^{J/2}\Gamma(\frac{J}{2})} \int_0^\infty P(\chi^2_{j+1} \ge (cq_J/\sqrt{v}+\tau/\sigma)^2)(q_J^2)^{J/2-1} e^{-q_J^2/2} dq_J^2 \tag{6.A.1}$$

Alternatively, if $\tau > 0$,

$$E\left[I_{(-\infty,cq_J/\sqrt{v}+\tau/\sigma)}(u_1)u_1^j\right] = \frac{1}{2^{J/2}\Gamma(\frac{J}{2})} \left[\int_0^\infty P(\chi^2_{j+1} \ge (cq_J/\sqrt{v}+\tau/\sigma)^2)(q_J^2)^{J/2-1} \right.$$
$$\left. \times\ e^{-q_J^2/2}\, dq_J^2 + I(t) \int_0^{\tau^2/c^2} P(\chi^2_{j+1} < (cq_J/\sqrt{v}+\tau/\sigma)^2)\ q_J^{2\,v/2-1}\, e^{-q_J^2/2}\, dq_J^2 \right], \tag{6.A.2}$$

where $I(t) = 0$ if t is odd, 1 otherwise.

Theorem 6.1 corresponds to a special case of Theorem 4.2 with $\lambda = \vartheta = 0$. Corollaries 6.3 to 6.7 follow by setting $j = 0, 1, \ldots, 4$ respectively.

Lemma 6.1 :

$\rho(\sigma^{**2}, \sigma^2)$ approaches $\rho(\sigma^{*2}, \sigma^2)$ as τ approaches infinity.

Proof :

From the text, as $\tau \to \infty$, $\rho(\sigma^{**2}, \sigma^2)$ becomes asympototic to

$$\rho(\tilde{\sigma}^2, \sigma^2) + \Big\{(\tau/\sigma)^2(n+\delta)^2\Big[(\tau/\sigma)^2 - 2(n+\gamma)\Big] + 2v(n+\delta)\Big[(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)(n+\gamma)\Big] - v(v+2)(\gamma-\delta)(2n + \gamma + \delta)\Big\}/\Big[(n+\delta)(n+\gamma)\Big]^2 + \Big\{2\Big[v + 3(\tau/\sigma)^2 - (n+\gamma)\Big] + 3\Big\}/(n+\gamma)^2 \quad . \qquad (6.A.3)$$

Now, when $\delta = \gamma = 0$, (6.A.3) becomes

$$\Big\{(2v + k^2) + (\tau/\sigma)^4 - 2n(\tau/\sigma)^2 + 2v(\tau/\sigma)^2 + \Big[2v + 6(\tau/\sigma)^2 - 2n + 3\Big]\Big\}/n^2$$

$$= \Big\{3 + k^2 + 6(\tau^2/\sigma^2) + \tau^4/\sigma^4 + 2v - 2k - 2k(\tau^2/\sigma^2)\Big\}/n^2,$$

which is the risk of σ^{*2}_{ML}. When $\delta = -k$ and $\gamma = 1-k$, (6.A.3) reduces to

$$2/v + \Big\{(\tau/\sigma)^2v^2\Big[(\tau/\sigma)^2 - 2(v+1)\Big] + 2v^2\Big[(\tau/\sigma)^2v + v + 1\Big] - v(v+2)(2n - k + 1 - k)\Big\}/(v(v+1))^2 + \Big\{2\Big[v + 3(\tau/\sigma)^2 - (v+1)\Big] + 3\Big\}/(v+1)^2$$

$$= \Big[(\tau/\sigma)^4 + 4(\tau/\sigma)^2\Big]/(v+1)^2 + (2v^2 + 2v)/(v(v+1)^2)$$

$$= \Big[2 + 4(\tau/\sigma)^2 + (\tau/\sigma)^4 + 2v\Big]/(v+1)^2$$

which is the risk of σ^{*2}_{LS}. Finally, when $\delta = -k+2$ and $\gamma = -k+3$, (6.A.3) collapses to

$$(2v + 4)/(v+2)^2 + \Big\{(\tau/\sigma)^2(v+2)^2\Big[(\tau/\sigma)^2 - 2(v+3)\Big] + 2v(v+2)\Big[(\tau/\sigma)^2(v+2) + (v+3)\Big] - v(v+2)(2v+5)\Big\}/((v+2)(v+3))^2 + \Big\{2\Big[v + 3(\tau/\sigma)^2 - (v+3)\Big] + 3\Big\}/(v+3)^2$$

$$= \Big[(\tau/\sigma)^4 - 3\Big]/(v+3)^2 + \Big[2(v+3)^2 + 2v(v+3) - v(2v+5)\Big]/((v+2)(v+3)^2)$$

$$= \Big[(\tau/\sigma)^4 - 3\Big]/(v+3)^2 + (18 + 13v + 2v^2)/((v+2)(v+3)^2)$$

$$= \Big[(\tau/\sigma)^4 - 3\Big]/(v+3)^2 + \Big[(v+2)(2v+9)\Big]/((v+2)(v+3)^2)$$

$$= \left[6 + 2v + (\tau^4/\sigma^4)\right]/(v+3)^2,$$

which is the risk of σ^{*2}_{MM}.

Lemma 6.2 :

$\rho(\hat{\sigma}^2, \sigma^2)$ approaches $\rho(\tilde{\sigma}^2, \sigma^2)$ as $|\tau|$ approaches infinity.

Proof :

As $\tau \to -\infty$, $\bar{E}_{IL} \to 0$. Accordingly, the risk of $\hat{\sigma}^2$ (as given in (6.37)) approaches $\rho(\sigma^{**2}, \sigma^2)$. However, as noted earlier in the text, when $\tau \to -\infty$, $\rho(\sigma^{**2}, \sigma^2) \to \rho(\tilde{\sigma}^2, \sigma^2)$. Hence $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\tilde{\sigma}^2, \sigma^2)$ as $\tau \to -\infty$.

When $\tau \to \infty$, $\bar{E}_{IL} \to 0$ and $\bar{G}_{IL} \to 1$. Accordingly, $\rho(\hat{\sigma}^2, \sigma^2)$ approaches

$$\rho(\sigma^{**2}, \sigma^2) - \left\{(\tau/\sigma)^2(n+\delta)^2\left[(\tau/\sigma)^2 - 2(n+\gamma)\right] + 2v(n+\delta)\left[(\tau/\sigma)^2(n+\delta) + (\gamma-\delta)(n+\gamma)\right] - v(v+2)(\gamma-\delta)(2n + \gamma + \delta)\right\}/\left[(n+\delta)(n+\gamma)\right]^2 - \left\{2\left[v + 3(\tau/\sigma)^2 - (n+\gamma)\right] + 3\right\}/(n+\gamma)^2 \tag{6.A.4}$$

as $\tau \to \infty$. Substituting the expression for $\rho(\sigma^{**2}, \sigma^2)$ when $\tau \to \infty$, as given in (6.A.3), into (6.A.4), (6.A.4) reduces to the risk of $\tilde{\sigma}^2$. Hence $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\tilde{\sigma}^2, \sigma^2)$ as $\tau \to \infty$.

APPENDIX 6B

Figure 6.1

Relative risk functions of $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML}, σ^{**2}_{ML} and $\hat{\sigma}^2_{ML}$ for n = 20 and k = 5

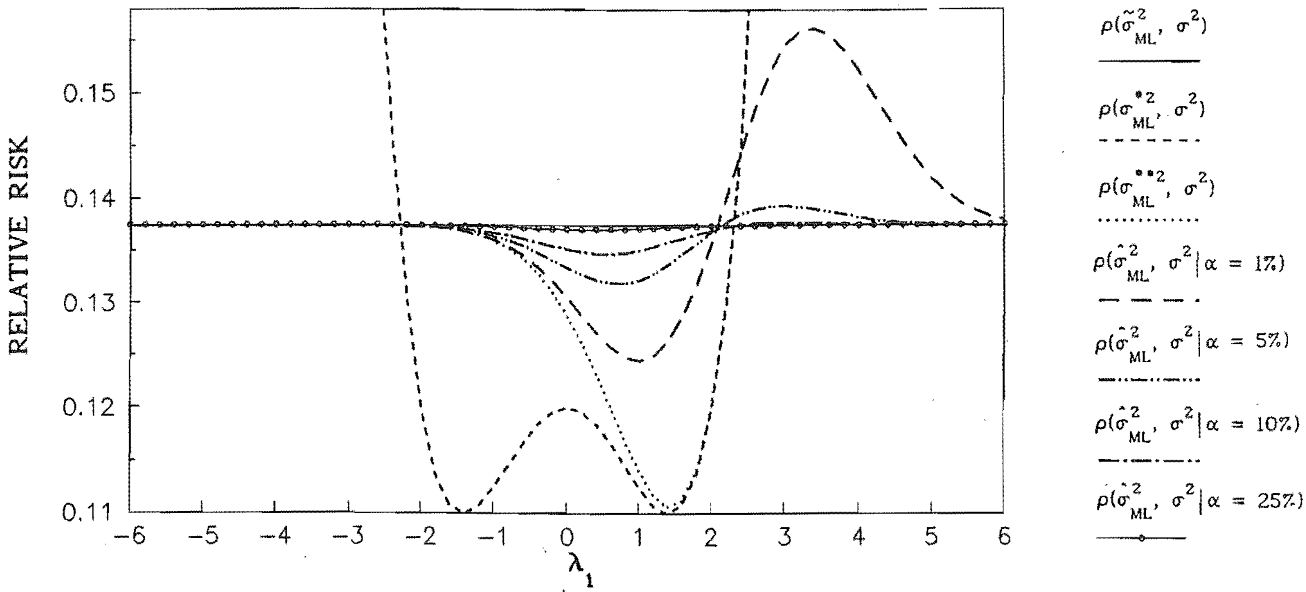


Figure 6.2

Relative risk functions of $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML}, σ^{**2}_{ML} and $\hat{\sigma}^2_{ML}$ for n = 30 and k = 2

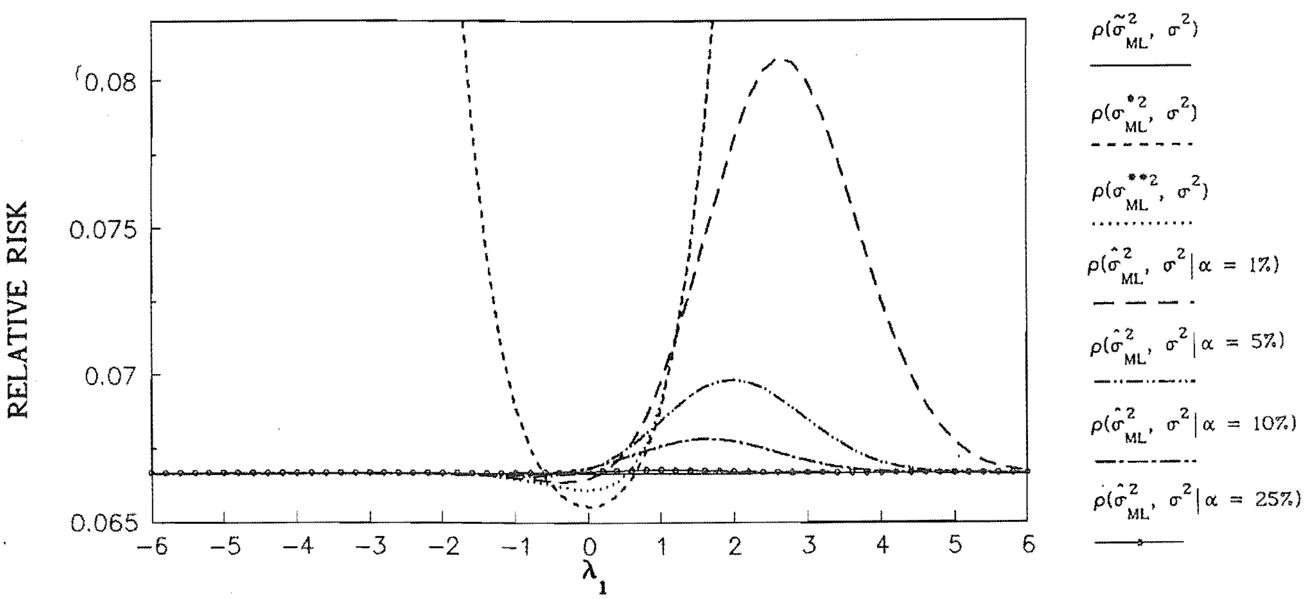


Figure 6.3

Relative risk functions of $\tilde{\sigma}^2_{LS}$, σ^{*2}_{LS}, σ^{**2}_{LS} and $\hat{\sigma}^2_{LS}$ for n = 20 and k = 5

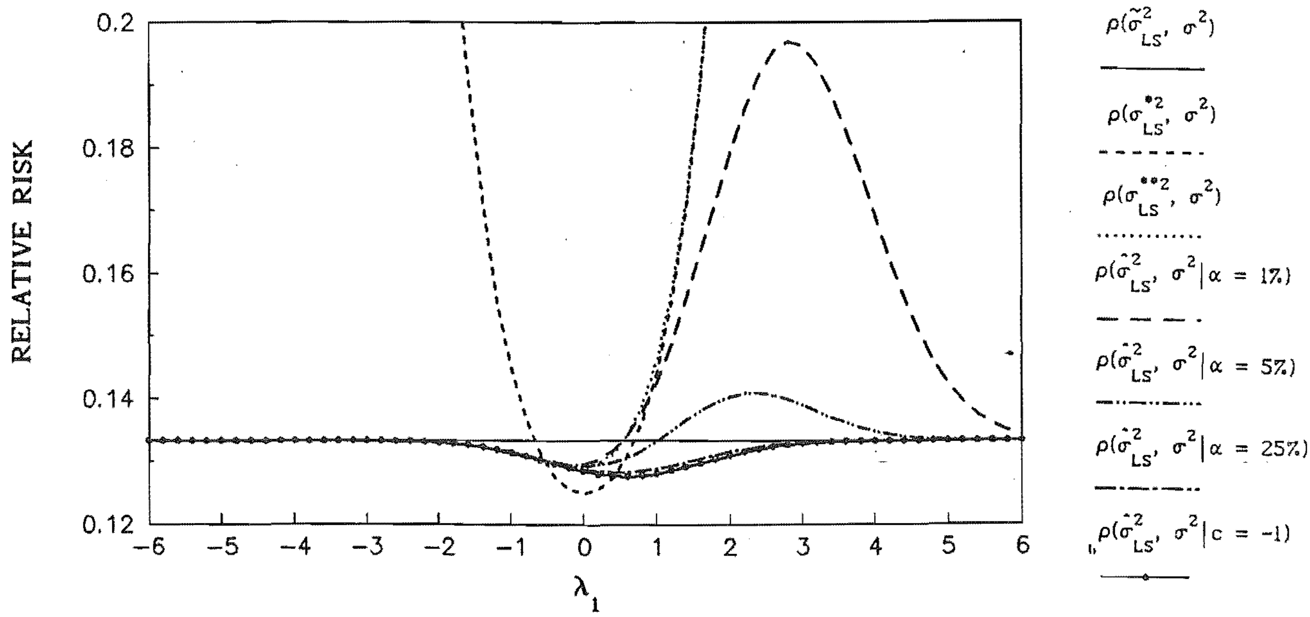


Figure 6.4

Relative risk functions of $\tilde{\sigma}^2_{LS}$, σ^{*2}_{LS}, σ^{**2}_{LS} and $\hat{\sigma}^2_{LS}$ for n = 30 and k = 2

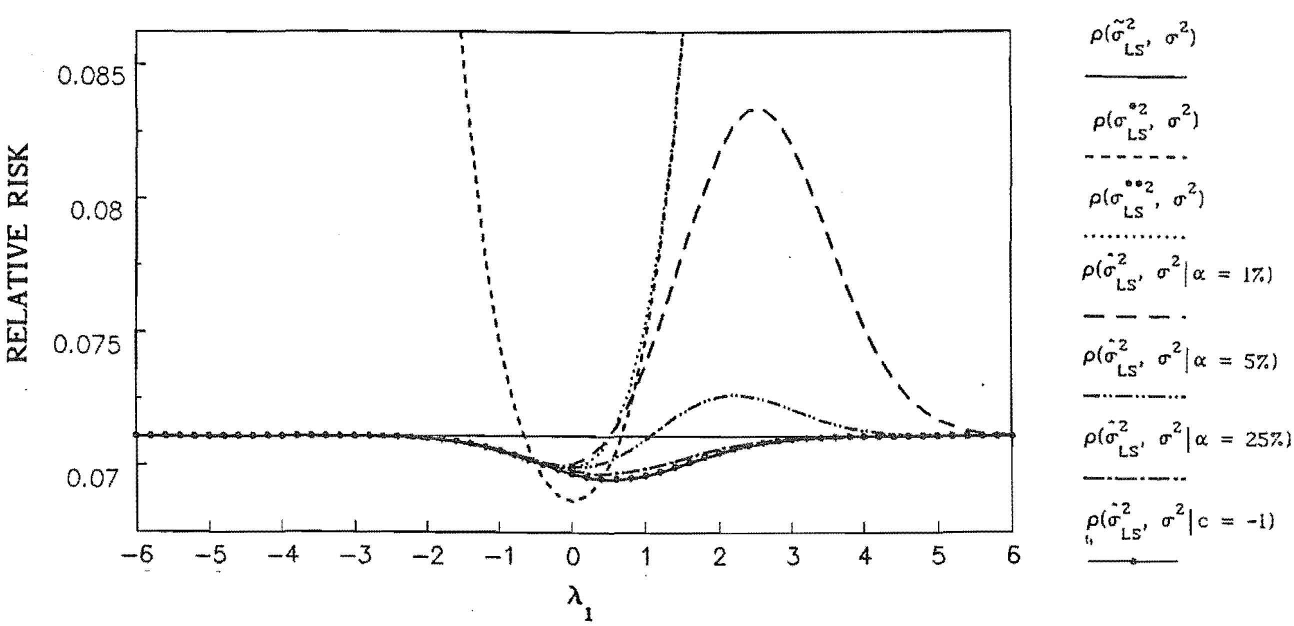


Figure 6.5

Relative risk functions of $\tilde{\sigma}^2_{MM}$, σ^{*2}_{MM}, σ^{**2}_{MM} and $\hat{\sigma}^2_{MM}$ for n = 20 and k = 5

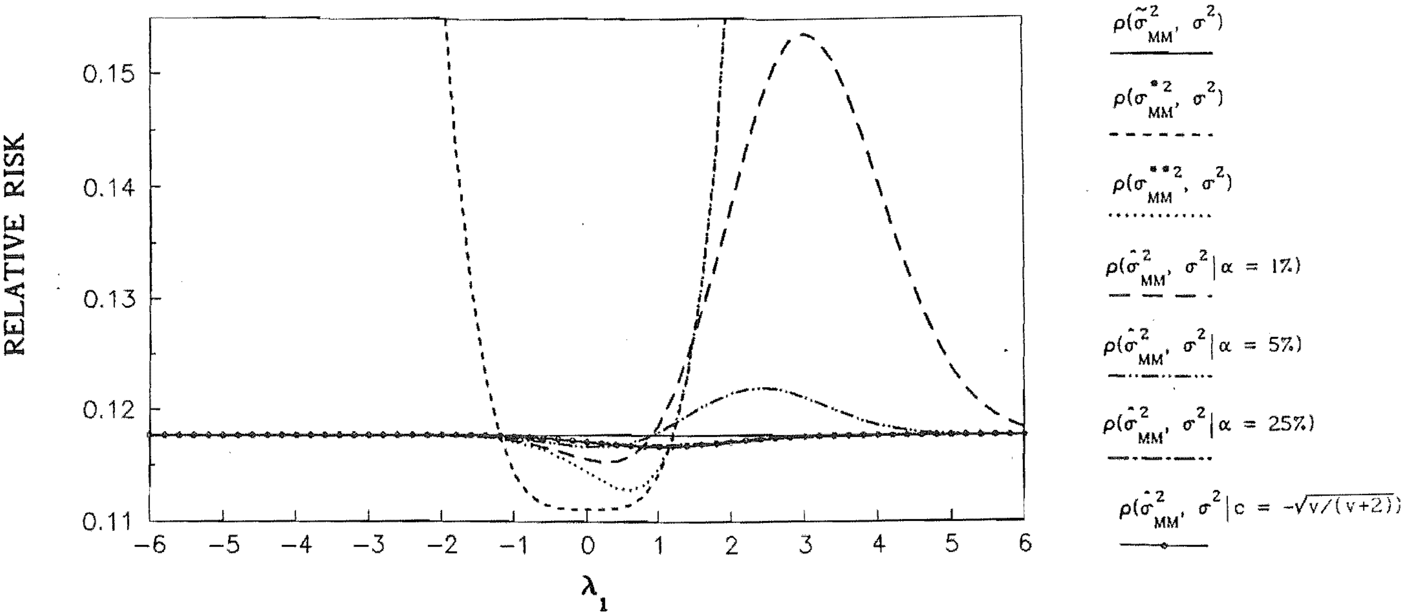
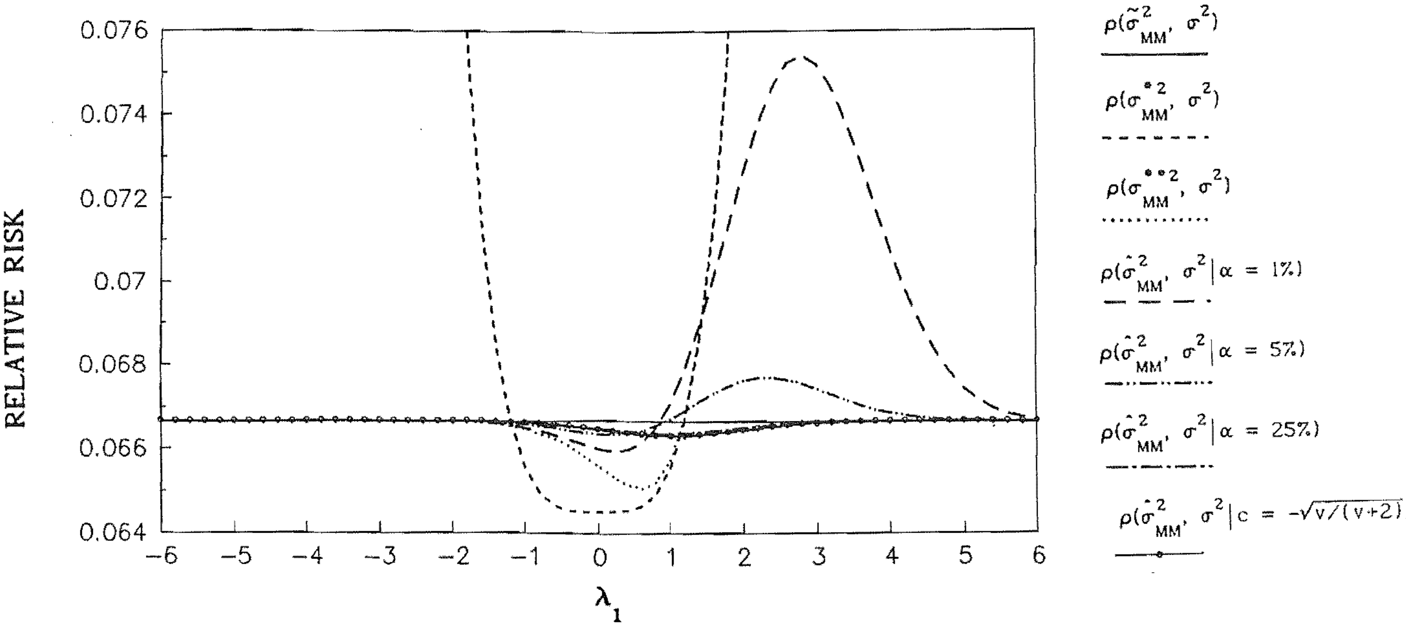


Figure 6.6

Relative risk functions of $\tilde{\sigma}^2_{MM}$, σ^{*2}_{MM}, σ^{**2}_{MM} and $\hat{\sigma}^2_{MM}$ for n = 30 and k = 2



CHAPTER SEVEN
THE SAMPLING PERFORMANCE OF THE INEQUALITY RESTRICTED AND PRE-TEST ESTIMATORS IN AN UNDERFITTED MODEL

7.1 INTRODUCTION

Having considered in the previous chapter the properties of the inequality restricted and pre-test estimators for the scale parameter in the standard linear model, we now examine the effects of underfitting the model on the properties of these estimators. The consequences of such a model mis-specification for the properties of the estimators of the prediction vector have been discussed in Chapter 4.

Within the model framework established in Chapter 4, we derive and numerically evaluate the risk functions under quadratic loss for the general family of estimators of the error variance considered in Chapter 6. The members of this family are the maximum likelihood, least squares and the minimum mean squared error component estimators. We assume that the researcher considers model (4.2), which is mis-specified through the exclusion of relevant regressors. As is well known, when the model is underfitted, the quadratic form $\tilde{e}'\tilde{e}/\sigma^2$ no longer has a central chi-square distribution as stated in the preceding chapter, but it is distributed instead as a non-central chi-square random variable. The ratio $e^{*\prime}e^{*}/\sigma^2$, on the other hand, is still distributed as a non-central chi-square variate as in the well specified model, but the non-centrality parameter associated with its distribution is now dependent on both the constraint and the model specification errors. In particular, the non-centrality parameter does not generally collapse to zero when the restriction holds as a strict equality.

Given these facts and also the fact that both the inequality restricted and pre-test estimators are, in one way or another, stochastic mixtures of the unrestricted and equality restricted estimators, underfitting the model is therefore likely to have implications for the properties of the inequality restricted and pre-test estimators. As in the case of estimating the prediction vector, this type of model mis-specification affects the properties of the latter estimator not only through the component estimators, but also through the test statistic.

This chapter is organised as follows. In the next section, we shall briefly re-state the statistical model presented in Chapter 4 and derive the exact finite sample risk functions of the inequality restricted estimators for the scale parameter. The risk functions of the corresponding inequality pre-test estimators are examined in Section 7.3. This extends the results given in Chapter 6 by considering the risk functions of these estimators when the model is underfitted. Some concluding remarks are given in Section 7.4.

7.2 THE STATISTICAL MODEL AND RISK FUNCTIONS OF THE ESTIMATORS FOR THE SCALE PARAMETER

As in Chapter 4, we assume that the true data generating process is

$$y = X\beta + Z\eta + \varepsilon \quad ; \qquad \varepsilon \sim N(0, \sigma^2 I) \tag{7.1}$$

where y, X, β, Z, η and ε are defined as previously. We assume, however, that the set of regressors Z is mistakenly omitted from the model. The fitted model is therefore

$$y = X\beta + \mu \quad ; \qquad \mu \sim N(Z\eta, \sigma^2 I) \quad . \tag{7.2}$$

As the researcher is unaware of the mis-specification in the model, $E(\varepsilon)$ is incorrectly assumed to be zero. As in the previous chapters, (7.1) and (7.2) can be reparameterized into the orthonormal models

$$y = H\theta + B\pi + \varepsilon \tag{7.3}$$

and

$$y = H\theta + \mu \tag{7.4}$$

respectively. The prior information available to the researcher is represented by $C'\beta \geq r$ in terms of the original model, or $\theta_1 \geq r_0$ in terms of the reparameterized model, where r_0 is a (positive) scalar multiple of r and θ_1 is the first element in θ.

Under (7.4), the unrestricted estimator (UE) and equality restricted estimator (ERE) of σ^2 are $\tilde{\sigma}^2 = \tilde{e}'\tilde{e}/(n+\delta)$ and $\sigma^{*2} = e^{*\prime}e^*/(n+\gamma)$ respectively, where $\tilde{e}$ and e^* are the vectors of residuals corresponding to the use of the UE and ERE of θ in model (7.4) respectively. The values of γ and δ are defined as in Chapter 6 for the least squares, maximum likelihood and minimum mean squared error members of the general family of component estimators.

It is straightforward to show that under the stated assumptions, $(n+\delta)\tilde{\sigma}^2/\sigma^2 \sim \chi'^2_{(v;\lambda_2)}$ and $(n+\gamma)\sigma^{*2}/\sigma^2 \sim \chi'^2_{(v+1;\lambda_1^2+\lambda_2)}$, where $\lambda_1 = (\tau - \xi)/(\sqrt{2}\sigma)$, $\lambda_2 = \pi'B'(I-HH')B\pi/(2\sigma^2)$, $\tau = r_0 - \theta_1$ and $\xi = (H'B\pi)_1$ is the first element of $H'B\pi$. Using the moments of non-central Chi-square variables,

$$E(\tilde{\sigma}^2) = \sigma^2(v + 2\lambda_2)/(n+\delta), \tag{7.5}$$

$$\mathrm{var}(\tilde{\sigma}^2) = 2\sigma^4(v + 4\lambda_2)/(n+\delta)^2, \tag{7.6}$$

$$E(\sigma^{*2}) = \sigma^2(1 + v + 2(\lambda_1^2 + \lambda_2))/(n+\gamma), \tag{7.7}$$

and

$$\mathrm{var}(\sigma^{*2}) = 2\sigma^4(1 + v + 4(\lambda_1^2 + \lambda_2))/(n+\gamma)^2. \tag{7.8}$$

As σ^2 is a scalar, the risk of any estimator of σ^2 is simply its relative mean squared error. Therefore,

$$\rho(\tilde{\sigma}^2, \sigma^2) = [2(v + 4\lambda_2) + (v + 2\lambda_2 - (n+\delta))^2]/(n+\delta)^2 \tag{7.9}$$

and

$$\rho(\sigma^{*2}, \sigma^2) = \Big[2[1 + v + 4(\lambda_1^2 + \lambda_2)] + [1 - k + 2(\lambda_1^2 + \lambda_2) - \gamma]^2\Big]/(n+\gamma)^2. \qquad (7.10)$$

As the researcher is unaware of the possible mis-specification of the model, its existence has no bearing on the specification of an estimator. Therefore, as in Chapter 6, the inequality restricted estimator (IRE) of σ^2 is

$$\sigma^{**2} = \begin{cases} \tilde{\sigma}^2 & \text{if } \tilde{\theta} \ge r \\ \sigma^{*2} & \text{if } \tilde{\theta} < r \end{cases} . \qquad (7.11)$$

So, the risk of σ^{**2} may be expressed as :

$$\rho(\sigma^{**2}, \sigma^2) = \rho(\tilde{\sigma}^2, \sigma^2) + E\Big\{I_{(-\infty,\tau/\sigma)}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big] \times\Big[2(\tilde{\sigma}^2-\sigma^2) + \Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \quad . \qquad (7.12)$$

The evaluation of this risk involves the evaluations of $E\Big[\tilde{\sigma}^i I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\Big]$, $i = 2, 4$ and $j = 0, 1, 2, 3, 4$. Given that u_1 is distributed independently[1] of $\tilde{\sigma}^2$, $E\Big[\tilde{\sigma}^i I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\Big]$ can be written as $E(\tilde{\sigma}^i)$ $E\Big[I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\Big]$.

Now, from (7.5), $E(\tilde{\sigma}^2) = \sigma^2(v + 2\lambda_2)/(n+\delta)$. Furthermore, $E(\tilde{\sigma}^4) = \sigma^4 E(\chi'^2_{(v;\lambda_2)}\chi'^2_{(v;\lambda_2)})/(n+\delta)^2$. Applying Lemma 2 of Judge and Bock (1978, pp. 322), or Lemma 1 of Clarke *et al.* (1987a), we can show[2]

$$E(\chi'^2_{(v;\lambda_2)}\chi'^2_{(v;\lambda_2)}) = E[(\chi'^2_{(v+2;\lambda_2)})]\mathrm{tr}(I_v) + E[(\chi'^2_{(v+4;\lambda_2)})]2\lambda_2. \qquad (7.13)$$

Now, $\mathrm{tr}(I_v) = v$. $E[(\chi'^2_{(v+2;\lambda_2)})] = (v + 2 + 2\lambda_2)$ and $E[(\chi'^2_{(v+4;\lambda_2)})] = (v + 4 + 2\lambda_2)$. It then follows that

---

[1] The assumption of an underfitted model does not affect this result. The joint distribution of u and $\tilde{\sigma}^2$ is still equal to the product of their marginal distributions.

[2] An alternative way of showing this result is to make use of the fact that $\mathrm{var}(\tilde{\sigma}^2) = E\Big[\tilde{\sigma}^2 - E(\tilde{\sigma}^2)\Big]^2$.

$$
\begin{aligned}
E(\tilde{\sigma}^4) &= \sigma^4\Big[(v + 2 + 2\lambda_2)v + (v + 4 + 2\lambda_2)2\lambda_2\Big]/(n+\delta)^2 \\
&= \sigma^4\Big[v(v + 2) + 4\lambda_2(v + \lambda_2 + 2)\Big]/(n+\delta)^2. \qquad (7.14)
\end{aligned}
$$

In evaluating $E\Big[I_{(-\infty,\tau/\sigma)}(u_1)u_1^j\Big]$, $j = 0, 1, 2, 3, 4$, we use Theorem 4.1 which is given and proved in Appendix 4A of Chapter 4. The results for the special cases of $j = 0$, 1 and 2 are given in Corollaries 4.1 to 4.3 in Chapter 4.

From Theorem 4.1 if we set $j = 3$, we obtain

<u>Corollary 7.1</u>

$$
E\Big[I_{(-\infty,\tau/\sigma)}(u_1)u_1^3\Big] = \begin{cases} \dfrac{\xi^3}{2\sigma^3}P(\chi_1^2 \geq 2\lambda_1^2) - \dfrac{3\xi^2}{\sigma^2\sqrt{2\pi}}P(\chi_2^2 \geq 2\lambda_1^2) + \dfrac{3\xi}{2\sigma}P(\chi_3^2 \geq 2\lambda_1^2) \\ \qquad \times \sqrt{\dfrac{2}{\pi}}P(\chi_4^2 \geq 2\lambda_1^2) \\ \\ \dfrac{\xi^3}{\sigma^3} - \dfrac{\xi^3}{2\sigma^3}P(\chi_1^2 \geq 2\lambda_1^2) - \dfrac{3\xi^2}{\sigma^2\sqrt{2\pi}}P(\chi_2^2 \geq 2\lambda_1^2) + \dfrac{3\xi}{\sigma} \\ \qquad - \dfrac{3}{2}P(\chi_3^2 \geq 2\lambda_1^2) - \sqrt{\dfrac{2}{\pi}}P(\chi_4^2 \geq 2\lambda_1^2) \end{cases} \qquad (7.15)
$$

Further, by using Theorem 4.1 and setting $j = 4$, we have

<u>Corollary 7.2</u>

$$
E\Big[I_{(-\infty,\tau/\sigma)}(u_1)u_1^4\Big] = \begin{cases} \dfrac{\xi^4}{2\sigma^4}P(\chi_1^2 \geq 2\lambda_1^2) - \dfrac{4\xi^3}{\sigma^3\sqrt{2\pi}}P(\chi_2^2 \geq 2\lambda_1^2) + \dfrac{3\xi^2}{\sigma^2}P(\chi_3^2 \geq 2\lambda_1^2) \\ \qquad - \dfrac{4\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}P(\chi_4^2 \geq 2\lambda_1^2) + \dfrac{3}{2}P(\chi_5^2 \geq 2\lambda_1^2) \\ \\ \dfrac{\xi^4}{\sigma^4} - \dfrac{\xi^4}{2\sigma^4}P(\chi_1^2 \geq 2\lambda_1^2) - \dfrac{4\xi^3}{\sigma^3\sqrt{2\pi}}P(\chi_2^2 \geq 2\lambda_1^2) + \dfrac{6\xi^2}{\sigma^2} \\ \qquad - \dfrac{3\xi^2}{\sigma^2}P(\chi_3^2 \geq 2\lambda_1^2) - \dfrac{4\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}P(\chi_4^2 \geq 2\lambda_1^2) \\ \qquad + 3 - \dfrac{3}{2}P(\chi_5^2 \geq 2\lambda_1^2) \end{cases} \qquad (7.16)
$$

Making use of these corollaries along with results (7.5) and (7.14), and after performing some tedious manipulations,[3] we show that the risk of σ^{**2}, when $\lambda_1 \leq 0$, is

$$\rho(\sigma^{**2}, \sigma^2) = \rho(\tilde{\sigma}^2, \sigma^2) + \Big\{2\lambda_1^2(n+\delta)^2\Big[2\lambda_1^2 - 2(n+\gamma)\Big] + 2(v+2\lambda_2)(n+\delta)\Big[2\lambda_1^2(n+\delta) + (n+\gamma)(\gamma-\delta)\Big] - \Big[v(v+2) + 4\lambda_2^2 + 4\lambda_2 v + 8\lambda_2\Big](\gamma-\delta)(2n + \gamma + \delta)\Big\} P_1/2\Big[(n+\gamma)(n+\delta)\Big]^2 + \Big[8\lambda_1^3 + 4\lambda_1 v + 8\lambda_1\lambda_2 - 4\lambda_1(n+\gamma)\Big]P_2/\Big[\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + \Big[6\lambda_1^2 + v - (n+\gamma) + 2\lambda_2\Big]P_3/(n+\gamma)^2 + 8\lambda_1 P_4/\Big[\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + 3P_5/(2(n+\gamma)^2) \qquad (7.17)$$

Alternatively, when $\lambda_1 > 0$, the risk of σ^{**2} is

$$\rho(\sigma^{**2}, \sigma^2) = \rho(\tilde{\sigma}^2, \sigma^2) + \Big\{2\lambda_1^2(n+\delta)^2\Big[2\lambda_1^2 - 2(n+\gamma)\Big] + 2(v+2\lambda_2)(n+\delta)\Big[2\lambda_1^2(n+\delta) + (n+\gamma)(\gamma-\delta)\Big] - \Big[v(v+2) + 4\lambda_2^2 + 4\lambda_2 v + 8\lambda_2\Big](\gamma-\delta)(2n + \gamma + \delta)\Big\} \times \Big[1 - P_1/2\Big]/\Big[(n+\gamma)(n+\delta)\Big]^2 + \Big[8\lambda_1^3 + 4\lambda_1 v + 8\lambda_1\lambda_2 - 4\lambda_1(n+\gamma)\Big]P_2 /\Big[\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + \Big[6\lambda_1^2 + v - (n+\gamma) + 2\lambda_2\Big](2-P_3)/(n+\gamma)^2 + 8\lambda_1 P_4 /\Big[\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + (6 - 3P_5)/(2(n+\gamma)^2) \quad , \qquad (7.18)$$

where $P_I = P(\chi_I^2 \geq 2\lambda_1^2)$, $I = 1,..,5$.

When there are no omitted variables, $\lambda_2 = 0$, $\lambda_1 = \tau/(\sqrt{2}\sigma)$ and (7.17) and (7.18) collapse to the expressions given in Chapter 6 for the case in which the model is well specified.

Using the Convergence Theorem given in Judge and Yancey (1986, p. 77), we can show that for any given λ_2, $\lambda_1 P(\chi_I^2 \geq 2\lambda_1^2) \to 0$ as $\lambda_1 \to \pm\infty$. Accordingly, as

---

[3] Details are available upon request.

$\lambda_1 \rightarrow -\infty$, $\rho(\sigma^{**2}, \sigma^2) \rightarrow \rho(\tilde{\sigma}^2, \sigma^2)$. Similarly, as $\lambda_1 \rightarrow \infty$, $\rho(\sigma^{**2}, \sigma^2) \rightarrow \rho(\sigma^{*2}, \sigma^2)$. We use a similar approach as outlined in the last chapter. These results concur with those given in the previous chapter when the model is well specified, and are consistent with the results when estimating E(y).

We have numerically evaluated the risk of σ^{**2} for n = 20, 30, 40, 50, 80, k = 2, 5, 10, 15, ..., n-5, $\lambda_1 \in [-10, 10]$ and various choices of λ_2 in ways similar to those described in Chapter 6. Some representative diagrams are given in Appendix 7B. If the component estimator is LS or MM, then our numerical results illustrate that the inequality $\rho(\sigma^{**2}, \sigma^2) \leq \rho(\tilde{\sigma}^2, \sigma^2)$ always holds when $\lambda_1 \leq 0$, or nearly so. $\rho(\sigma^{**2}, \sigma^2)$ intersects $\rho(\tilde{\sigma}^2, \sigma^2)$ when $\lambda_1 > 0$ and increases without bound thereafter. The biggest risk gain from using σ^{**2}_{LS} or σ^{**2}_{MM} tends to occur when $\lambda_1 < 0$. This constrasts with the corresponding result observed when estimating the prediction vector, where the biggest risk gain of using $X\beta^{**}$ over $X\tilde{\beta}$ always occurs at the origin (*i.e.*, $\lambda_1 = 0$).

However, analgous to the case when estimating E(y), the results reported here for $\lambda_1 \leq 0$ do not necessarily imply that σ^{**2} is risk superior to $\tilde{\sigma}^2$ when the prior information is correct. When the model is underfitted, both τ, the surplus variable, and ξ, the model specification error enter into the definition of λ_1. A negative τ, and a sufficiently large ξ will result in a value of λ_1 such that $\rho(\sigma^{**2}, \sigma^2) > \rho(\tilde{\sigma}^2, \sigma^2)$. The use of valid prior information therefore does not guarantee a reduction in risk when the model is underfitted. This is consistent with the results obtained when estimating E(y), and also for the case in which the prior information exists in an exact equality form (Giles and Clarke (1989)).

The situation changes dramatically when the component estimator is ML. With a relatively large λ_2, σ^{**2}_{ML} can be inferior to the UE over the entire λ_1 space. In fact, $\tilde{\sigma}^2_{ML}$ can strictly dominate both σ^{**2}_{ML} and σ^{*2}_{ML} when λ_2 is sufficiently large (see Figures 7.1 and 7.2). This does not occur if the component estimator is the LS or MM estimator instead.

7.3 THE RISK OF $\hat{\sigma}^2$

Under the framework of model (7.2), the inequality pre-test estimator (IPTE) of σ^2 is

$$\hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2 & \text{if } t'' \le c \\ \sigma^{**2} & \text{if } t'' > c \end{cases} = I_{(-\infty,c]}(t'')\tilde{\sigma}^2 + I_{(c,\infty)}(t'')\sigma^{**2} \quad , \tag{7.19}$$

where $t'' = \sqrt{v}(\tilde{\theta}_1 - r_0)\tilde{\sigma}^{-1}/\sqrt{n+\delta}$ has a non-central t distribution with non-centrality parameters λ_1^2 and λ_2, and c is the size - α critical value for the *central* t variate with v degrees of freedom as defined in Chapter 4.

Analgous to Chapter 6, we rewrite (7.19) as

$$\hat{\sigma}^2 = \sigma^{**2} - I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\left[\left((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\right)/(n+\delta)\right], \tag{7.20}$$

where $c' = c\sqrt{(n+\delta)/v}$, and the corresponding risk function is

$$\rho(\hat{\sigma}^2, \sigma^2) = \rho(\sigma^{**2}, \sigma^2) - E\left\{I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\left[\left((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\right)/(n+\gamma)\right]\right.$$
$$\left.\times\left[2(\tilde{\sigma}^2 - \sigma^2) + \left((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\right)/(n+\gamma)\right]\right\}/\sigma^4. \tag{7.21}$$

The evaluation of this risk involves the evaluations of $E\left[\tilde{\sigma}^i I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$, $i = 0, 2, 4$ and $j = 0, 1,..,4$. First, let us consider $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$. This expectation is evaluated using Theorem 4.2 stated and proved in Appendix 4A of Chapter 4. The results for the special cases of $j = 0, 1, 2$ are given by Corollaries 4.4 to 4.6 of Chapter 4. Now, by using Theorem 4.2 and setting $j = 3, 4$, we obtain the following corollaries :

Corollary 7.3

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^3\right]$$

$$= \begin{cases} \dfrac{\xi^3}{2\sigma^3}E_{1,v} - \dfrac{3\xi^2}{\sigma^2\sqrt{2\pi}}E_{2,v} + \dfrac{3\xi}{2\sigma}E_{3,v} - \sqrt{\dfrac{2}{\pi}}E_{4,v} & \text{if } \lambda_1 \le 0 \\ \dfrac{\xi^3}{2\sigma^3}E_{1,v} + \dfrac{\xi^3}{\sigma^2}G_{1,v} - \dfrac{3\xi^2}{\sigma^2\sqrt{2\pi}}E_{2,v} + \dfrac{3\xi}{2\sigma}E_{3,v} + \dfrac{3\xi}{\sigma}G_{3,v} - \sqrt{\dfrac{2}{\pi}}E_{4,v} & \text{if } \lambda_1 > 0 \end{cases}, \tag{7.22}$$

Corollary 7.4

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^4\right]$$

$$= \begin{cases} \dfrac{\xi^4}{2\sigma^4}E_{1,v} - \dfrac{4\xi^3}{\sigma^3\sqrt{2\pi}}E_{2,v} + \dfrac{3\xi^2}{\sigma^2}E_{3,v} - \dfrac{4\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}E_{4,v} + \dfrac{3}{2}E_{5,v} & \text{if } \lambda_1 \le 0 \\ \dfrac{\xi^4}{2\sigma^4}E_{1,v} + \dfrac{\xi^4}{\sigma^4}G_{1,v} - \dfrac{4\xi^3}{\sigma^3\sqrt{2\pi}}E_{2,v} + \dfrac{3\xi^2}{\sigma^2}E_{3,v} + \dfrac{6\xi^2}{\sigma^2}G_{3,v} - \dfrac{4\xi}{\sigma}\sqrt{\dfrac{2}{\pi}}E_{4,v} & \\ \qquad + \dfrac{3}{2}E_{5,v} - 3G_{5,v} & \text{if } \lambda_1 > 0 \end{cases}, \tag{7.23}$$

where, as in Chapter 4,

$$E_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty}\frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)}\int_0^{\infty} P(\chi_I^2 \ge (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2)(q_J^2)^{J/2+t-1}e^{-q_J^2/2}\,dq_J^2$$

and

$$G_{I,J} = e^{-\lambda_2}\sum_{t=0}^{\infty}\frac{\lambda_2^t}{t!\,2^{J/2+t}\Gamma(\frac{J}{2}+t)}\int_0^{2v\lambda_1^2/c^2} P(\chi_I^2 < (cq_J/\sqrt{v}+\sqrt{2}\lambda_1)^2)(q_J^2)^{J/2+t-1}e^{-q_J^2/2}\,dq_J^2,$$

where q_J^2 is a non-central Chi-Squared random variable with J degrees of freedom and $I = 1,\ldots,5$.

The use of these two corollaries, together with Corollaries 4.4 to 4.6 given in Chapter 4, facilitates the evaluation of $E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$, j =

0, 1,..5.

Second, let us consider $E\left[\tilde{\sigma}^2 I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$ and $E\left[\tilde{\sigma}^4 I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j\right]$. Again, as in Chapter 6, $\tilde{\sigma}^2$ and $I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j$ are not independently distributed. However, as $\tilde{\sigma}$ is non-negative by definition, for each $\tilde{\sigma}^2$, there is only one corresponding $\tilde{\sigma}$. Accordingly, $I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)u_1^j|u_1$ can be regarded as a function of $\tilde{\sigma}^2$, and the theorems of Judge and Bock (1978) or Clarke *et al.* (1987a) can be applied.

Recall that $(n+\delta)\tilde{\sigma}^2/\sigma^2$ has a non-central chi-square distribution with v degrees of freedom and non-centrality parameter λ_2. From Judge and Bock (1978, p. 322), we have

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\tilde{\sigma}^2\right] = \frac{\sigma^2}{n+\delta}\left\{vE\left[I_{(-\infty,\, cq_{v+2}/\sqrt{v}+\tau/\sigma)}(u_1)\right] + 2\lambda_2 \times E\left[I_{(-\infty,\, cq_{v+4}/\sqrt{v}+\tau/\sigma)}(u_1)\right]\right\}. \quad (7.24)$$

Using this result repeatedly, we obtain

$$E\left[I_{(-\infty,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\tilde{\sigma}^4\right]$$

$$= \frac{\sigma^4}{(n+\delta)^2}E\left[I_{(-\infty,\, cq_v/\sqrt{v}+\tau/\sigma)}(u_1)q_v^2q_v^2\right]$$

$$= \frac{\sigma^4}{(n+\delta)^2}\left\{vE\left(I_{(-\infty,\, cq_{v+2}/\sqrt{v}+\tau/\sigma)}(u_1)q_{v+2}^2\right) + 2\lambda_2E\left(I_{(-\infty,\, cq_{v+4}/\sqrt{v}+\tau/\sigma)}(u_1)q_{v+4}^2\right)\right\}$$

$$= \frac{\sigma^4}{(n+\delta)^2}\left\{v\left[(v+2)E\left(I_{(-\infty,\, cq_{v+4}/\sqrt{v}+\tau/\sigma)}(u_1)\right) + 2\lambda_2E\left(I_{(-\infty,\, cq_{v+6}/\sqrt{v}+\tau/\sigma)}(u_1)\right)\right]\right.$$
$$\left. + 2\lambda_2\left[(v+4)E\left(I_{(-\infty,\, cq_{v+6}/\sqrt{v}+\tau/\sigma)}(u_1)\right) + 2\lambda_2E\left(I_{(-\infty,\, cq_{v+8}/\sqrt{v}+\tau/\sigma)}(u_1)\right)\right]\right\}$$

$$= \frac{\sigma^4}{(n+\delta)^2}\left\{ v(v+2)E\left(I_{(-\infty,\, cq_{v+4}/\sqrt{v}+\tau/\sigma)}(u_1)\right) + 2v\lambda_2 E\left[I_{(-\infty,\, cq_{v+6}/\sqrt{v}+\tau/\sigma)}(u_1)\right)\right]\right.$$

$$\left. + 2\lambda_2(v+4)E\left[I_{(-\infty,\, cq_{v+6}/\sqrt{v}+\tau/\sigma)}(u_1)\right) + 4\lambda_2^2 E\left[I_{(-\infty,\, cq_{v+8}/\sqrt{v}+\tau/\sigma)}(u_1)\right)\right]\right\} . \tag{7.25}$$

Applying these results, along with Corollaries 4.4 to 4.6, 7.3 to 7.4, and after performing some tedious manipulations and re-arrangements, the risk of $\hat{\sigma}^2$, when $\lambda_1 \le 0$, is

$$\begin{aligned}
\rho(\hat{\sigma}^2, \sigma^2) = \; & \rho(\sigma^{**2}, \sigma^2) - 3E_{5,v}/(2(n+\gamma)) - 8\lambda_1 E_{4,v}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] + (n+\gamma-6\lambda_1^2) \\
& \times E_{3,v}/(n+\gamma)^2 + \left[4\lambda_1(n+\gamma)-8\lambda_1^3\right]E_{2,v}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] + \left[2\lambda_1^2(n+\gamma)-\lambda_1^4\right]E_{1,v} \\
& /(n+\gamma)^2 - vE_{3,v+2}/(n+\gamma)^2 - 4v\lambda_1 E_{2,v+2}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] - \left[v(n+\gamma)(\gamma-\delta)\right. \\
& \left. + 2v\lambda_1^2(n+\delta)\right]E_{1,v+2}/\left[(n+\delta)(n+\gamma)^2\right] - 2\lambda_2 E_{3,v+4}/(n+\gamma)^2 - 8\lambda_1\lambda_2 E_{2,v+4} \\
& /\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] + \left\{v(v+2)(\gamma-\delta)(2n+\gamma+\delta)/\left[2\left((n+\delta)(n+\gamma)\right)^2\right] - 4\lambda_1^2\lambda_2\right. \\
& \left. /(n+\gamma)^2 - 2(\gamma-\delta)\lambda_2/\left[(n+\delta)(n+\gamma)\right]\right\}E_{1,v+4} + (\gamma-\delta)(v+2)2\lambda_2(2n + \gamma + \delta) \\
& \times E_{1,v+6}/\left[(n+\delta)(n+\gamma)\right]^2 + (\gamma-\delta)2\lambda_2^2(2n + \gamma + \delta)E_{1,v+8}/\left[(n+\delta)(n+\gamma)\right]^2 .
\end{aligned} \tag{7.26}$$

Similarly, when $\lambda_1 > 0$, the risk of $\hat{\sigma}^2$ is

$$\begin{aligned}
\rho(\hat{\sigma}^2, \sigma^2) = \; & \rho(\sigma^{**2}, \sigma^2) - 3\left[E_{5,v}/2 + G_{5,v}\right]/(n+\gamma) - 8\lambda_1 E_{4,v}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] + (n+\gamma \\
& - 6\lambda_1^2)\left[E_{3,v} + 2G_{3,v}\right]/(n+\gamma)^2 + \left[4\lambda_1(n+\gamma)-8\lambda_1^3\right]E_{2,v}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] + \\
& \left[2\lambda_1^2(n+\gamma)-\lambda_1^4\right]\times\left[E_{1,v} + 2G_{1,v}\right]/(n+\gamma)^2 - v(E_{3,v+2} + 2G_{3,v+2})/(n+\gamma)^2 \\
& - 4v\lambda_1 E_{2,v+2}/\left[\Gamma(\tfrac{1}{2})(n+\gamma)^2\right] - \left[v(n+\gamma)(\gamma-\delta) + 2v\lambda_1^2(n+\delta)\right]\times\left[E_{1,v+2} + \right.
\end{aligned}$$

$$2G_{1,v+2}\Big]\Big/\Big[(n+\delta)(n+\gamma)^2\Big] - 2\lambda_2\Big[E_{3,v+4} + 2G_{3,v+4}\Big]\Big/(n+\gamma)^2 - 8\lambda_1\lambda_2E_{2,v+4}$$

$$\Big/\Big[\Gamma(\tfrac{1}{2})(n+\gamma)^2\Big] + \Big\{v(v+2)(\gamma-\delta)(2n+\gamma+\delta)\Big/\Big[2\Big((n+\delta)(n+\gamma)\Big)^2\Big] - 4\lambda_1^2\lambda_2$$

$$\Big/(n+\gamma)^2 - 2(\gamma-\delta)\lambda_2\Big/\Big[(n+\delta)(n+\gamma)\Big]\Big\}\times\Big[E_{1,v+4} + 2G_{1,v+4}\Big] + (\gamma-\delta)(v+2)$$

$$\times\ 2\lambda_2(2n + \gamma + \delta)\Big[E_{1,v+6} + 2G_{1,v+6}\Big]\Big/\Big[(n+\delta)(n+\gamma)\Big]^2 + (\gamma-\delta)2\lambda_2^2(2n +$$

$$\gamma + \delta) \times \Big[E_{1,v+8} + 2G_{1,v+8}\Big]\Big/\Big[(n+\delta)(n+\gamma)\Big]^2 \quad . \tag{7.27}$$

When there is no mis-specification in the model, $\lambda_2 = 0$, $\lambda_1 = \tau/(\sqrt{2}\sigma)$ and (7.26) - (7.27) reduce to the corresponding expressions given in Chapter 6.

From the definitions of E_{IL} and G_{IL}, we observe that for any given λ_2, as $\lambda_1 \to \pm\infty$, $E_{IL} \to 0$ while $G_{IL} \to 1$. On the other hand, for any finite λ_1 and non-zero c, G_{IL} and E_{IL} both approach zero as $\lambda_2 \to \infty$. Furthermore, E_{IL} approaches P_I while G_{IL} approaches $1 - P_I$ as $c \to 0^-$. Both E_{IL} and G_{IL} approach zero as $c \to -\infty$. Using these results, we can show that $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\tilde{\sigma}^2, \sigma^2)$ as $\lambda_1 \to \pm\infty$ (given λ_2)[4] and that $\rho(\hat{\sigma}^2, \sigma^2) \to \rho(\sigma^{**2}, \sigma^2)$ as $\lambda_2 \to \infty$ (given λ_1). $\rho(\hat{\sigma}^2, \sigma^2)$ approaches $\rho(\tilde{\sigma}^2, \sigma^2)$ as $c \to 0$ and approaches $\rho(\sigma^{**2}, \sigma^2)$ as $c \to -\infty$. This is analgous to the corresponding result when estimating E(y).

As in the previous section, we have numerically evaluated the risk functions using the same values of the arguments as discussed there. In addition, we consider the cases of α = 0.01, 0.05, 0.10, 0.25, 0.40. E_{IL} and G_{IL} are evaluated using the D01AJF subroutine from the NAG (1991) Subroutine library and subroutines FACTLN and GAMMQ from Press *et al.* (1986). These risks are also illustrated in Figures 7.1 to 7.6 in Appendix 7B (see pp. 186-188). We note from the numerical results that for a sufficiently large λ_2, $\tilde{\sigma}^2_{ML}$ can

[4] This can be done in a way similar to that given in Appendix 6A for the case where the model is properly specified.

uniformly dominate $\hat{\sigma}^2_{ML}$. This result is of no surprise given that the risk of $\tilde{\sigma}^2_{ML}$ can be smaller than that of σ^{**2}_{ML} over the entire λ_1 space when λ_2 is large.

By constrast, if the component estimator is LS or MM, the inequality $\rho(\sigma^{**2}, \sigma^2) \le \rho(\hat{\sigma}^2, \sigma^2) \le \rho(\tilde{\sigma}^2, \sigma^2)$ always holds when $\lambda_1 \le 0$ irrespective of the level of c. Regardless of the level of λ_2, there always exists a family of IPTE's which can simultaneously dominate *all* other estimators over certain regions in the parameter space. A sub-class of this family of IPTE's also strictly dominates the unrestricted estimator over the entire λ_1 range. Within this sub-class of IPTE's, it is found numerically that the estimator with $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM) has the minimum risk in the region where pre-testing is the best.[5] The minimum risk boundary is achieved by using σ^{**2} for $\lambda_1 \in (-\infty, \lambda_1^*]$ and $\hat{\sigma}^2|c = c^*$ for $\lambda_1 \in (\lambda_1^*, \infty)$, where $c^* = -1$ (for LS) or $-\sqrt{v/(v+2)}$ (for MM) and λ_1^* is that value of $\lambda_1 > 0$ for which $\rho(\sigma^{**2}, \sigma^2) = \rho(\hat{\sigma}^2, \sigma^2|c = c^*)$. When λ_2 is relatively small, the risk of the IPTE with $c = c^*$ is larger than the risk of $\hat{\sigma}^2$ with $c < c^*$ in the region $\lambda_1 < \lambda_1^*$.

Our numerical results also show that, depending on the choice of c, the risk gain from using σ^{**2} over $\hat{\sigma}^2$ in the region $\lambda_1 < \lambda_1^*$ can be very slight. More strikingly, it is found that when λ_2 is sufficiently large, there exist certain sizes of the pre-test, including that corresponding to c^*, such that the risk of the IPTE attains the risk of $\rho(\sigma^{**2}, \sigma^2)$ in the region $\lambda_1 \le \lambda_1^*$. Given our previous result that the pre-test estimator with $c = c^*$ achieves minimum risk in the rest of the parameter space, this finding implies that $\hat{\sigma}^2|c = c^*$ is a strictly dominating estimator when λ_2 is large.

In Appendix 7A, we prove that the risk of the pre-test estimator achieves stationary points at $c = 0, -\infty$ (for ML), $c = -1, -\infty$ (for LS) and $c = -\sqrt{v/(v+2)}$,

---

[5] We also prove in Appendix 7A that the risk of the IPTE achieves a stationary point at $c = -1$ (for LS), or $c = -\sqrt{v/(v+2)}$ (for MM).

$-\infty$ (for MM), which coincides with the results obtained when the *a priori* restriction holds as a strict equality, as is shown by Giles (1990).

7.4 CONCLUSIONS

In this chapter, we have derived and evaluated the risk functions of the inequality restricted and pre-test estimators for the error variance in a model which is underfitted. We have assumed that the prior information is in the form of a single linear inequality restriction imposed on the regression coefficient vector, and focused our attention on the maximum likelihood, least squares and minimum mean squared error component estimators.

One broad feature of our results is that underfitting the model complicates the risk properties of the estimators of σ^2 more than it complicates the properties of the predictive risk functions as examined in Chapter 4. In particular, it is found that underfitting a model can give rise to a strictly dominating estimator of the error variance. When the degree of model mis-specification is serious, the results here show that it is better to use the unrestricted estimator when the method of maximum likelihood is applied; to pre-test with a critical value of negative unity when using the least squares component estimator; or to pre-test with a critical value of $-\sqrt{v/(v+2)}$ when using the minimum mean squared error component estimator.

The question of the choice of optimal critical value is still to be answered for those cases where there exists no strictly dominating estimator. However, given that when the component estimator is LS or MM, the risk gain of using an estimator other than the pre-test estimator with $c = c^*$ is typically very slight when $\lambda_1 < \lambda_1^*$, and that the pre-test estimator with $c = c^*$ is the minimum risk estimator in the rest of parameter space, it is apparent that both the mini-max regret critical value and the optimal critical value according to the criterion of minimum average risk, will not be significantly different from c^*. The issue on an optimal choice of c is formally explored and discussed in the next chapter.

APPENDIX 7A

Lemma 7.1 :

$\rho(\hat{\sigma}^2,\ \sigma^2)$ achieves a stationary point at $c = 0,\ -\infty$ (for ML), $-1,\ -\infty$ (for LS), $-\sqrt{v/(v+2)},\ -\infty$ (for MM).

Proof :

From (7.21), the risk of the inequality pre-test estimator is

$$\rho(\hat{\sigma}^2,\ \sigma^2) = \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(-\infty,\,(c'\tilde{\sigma}+\tau)/\sigma)}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big] \times\Big[2(\tilde{\sigma}^2 - \sigma^2) + \Big((\sigma u_1-\tau)^2-\tilde{\sigma}^2(\gamma-\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4. \quad (7.A.1)$$

Using the fact that $\tilde{\sigma}^2 = \sigma^2 q_v^2/(n+\delta)$, where $q_v^2 \sim \chi'^2_{(v;\lambda_2)}$, (7.A.1) can be rewritten as :

$$\rho(\hat{\sigma}^2,\ \sigma^2) = \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(-\infty,\,(cq_v/\sqrt{v}+\tau/\sigma)}(u_1)\Big[\Big((\sigma u_1-\tau)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big] \times \Big[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma u_1-\tau)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \quad . \quad (7.A.2)$$

For the purposes of our analysis, we let $\omega = u_1 - \xi/\sigma$, a standard normal variable, and rewrite (7.A.2) as

$$\begin{aligned}\rho(\hat{\sigma}^2,\ \sigma^2) &= \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(-\infty,\,(cq_v/\sqrt{v}+(\tau-\xi)/\sigma)}(\omega)\Big[\Big((\sigma(\omega+\xi/\sigma)-\tau)^2-\sigma^2 \times(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big] \times \Big[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma(\omega+\xi/\sigma)-\tau)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \\ &= \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(-\infty,\,(cq_v/\sqrt{v}+\sqrt{2}\lambda_1)}(\omega)\Big[\Big((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2-\sigma^2 \times(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big] \times \Big[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \quad . \quad (7.A.3)\end{aligned}$$

First, let us consider the case when $cq_v/\sqrt{v} + \sqrt{2}\lambda_1 \le 0$. If we let $d = |cq_v/\sqrt{v} + \sqrt{2}\lambda_1| = -cq_v/\sqrt{v} - \sqrt{2}\lambda_1 \ge 0$, then (7.A.3) may be written as

$$\rho(\hat{\sigma}^2,\ \sigma^2) = \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(-\infty,-d)}(\omega)\Big[\Big((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)$$
$$/(n+\gamma)\Big] \times \Big[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2$$
$$-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \quad . \quad (7.A.4)$$

Proposition 1

For any given q_v, if $d \geq 0$, then

$$E\Big[I_{(-\infty,-d)}(\omega)\omega^j\Big] = \begin{cases} E\Big[I_{(d,\infty)}(\omega)\omega^j\Big] & \text{if } j \text{ is even or zero} \\ -E\Big[I_{(d,\infty)}(\omega)\omega^j\Big] & \text{if } j \text{ is odd} \end{cases} \quad (7.A.5)$$

Proof :

Given q_v, $E\Big[I_{(-\infty,-d)}(\omega)\omega^j\Big] = \displaystyle\int_{-\infty}^{-d} \frac{\omega^j e^{-\omega^2/2}}{\sqrt{2\pi}}\, d\omega.$

Now, $\dfrac{\omega^j e^{-\omega^2/2}}{\sqrt{2\pi}}$ is an odd function if j is odd, and an even function otherwise.

Using the properties of odd and even functions,

$$\int_{-\infty}^{-d} f(x)\ dx = \begin{cases} \displaystyle\int_{d}^{\infty} f(x)\ dx & \text{if } f \text{ is an even function} \\ -\displaystyle\int_{d}^{\infty} f(x)\ dx & \text{if } f \text{ is an odd function} \end{cases} .$$

Hence Proposition 1 follows directly. Q.E.D.

Using Proposition 1, for any given q_v,

$$E\Big[I_{(-\infty,-d)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^2\Big] = E\Big[I_{(d,\infty)}(\omega)(\sigma\omega + \sigma\sqrt{2}\lambda_1)^2\Big] \quad (7.A.6)$$

and

$$E\Big[I_{(-\infty,-d)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^4\Big] = E\Big[I_{(d,\infty)}(\omega)(\sigma\omega + \sigma\sqrt{2}\lambda_1)^4\Big]. \quad (7.A.7)$$

Therefore, (7.A.4) becomes

$$\rho(\hat{\sigma}^2,\ \sigma^2) = \rho(\sigma^{**2},\ \sigma^2) - E\Big\{I_{(d,\infty)}(\omega)\Big[\Big((\sigma\omega+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)$$
$$/(n+\gamma)\Big] \times \Big[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma\omega+\sigma\sqrt{2}\lambda_1)^2$$

$$-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \qquad . \qquad (7.A.8)$$

Now, for any given q_v,

$$E\left[I_{(d,\infty)}(\omega)\omega^j\right] = \int_d^\infty \frac{\omega^j e^{-\omega^2/2}}{\sqrt{2\pi}}\, d\omega.$$

$$= \frac{1}{2}\int_{d^2}^\infty \frac{\psi^{j/2}\ e^{-\psi/2}}{\sqrt{\psi}\ \sqrt{2}\Gamma(\frac{1}{2})}\, d\psi \qquad \text{if } \psi = \omega^2$$

$$= \frac{1}{2}\int_{d^2}^\infty \frac{\psi^{-1/2}\ e^{-\psi/2}}{\sqrt{2}\Gamma(\frac{1}{2})}\ \psi^{j/2}\ d\psi$$

$$= \frac{1}{2}\left[\int_0^\infty f(\psi)\ \psi^{j/2}\ d\psi - \int_0^{d^2} f(\psi)\ \psi^{j/2}\ d\psi\right] \quad , \qquad (7.A.9)$$

where $f(\psi) = \dfrac{\psi^{-1/2}\ e^{-\psi/2}}{\sqrt{2}\Gamma(\frac{1}{2})}$.

Hence (7.A.8) can be written as

$$\rho(\hat{\sigma}^2,\ \sigma^2) = \rho(\sigma^{**2},\ \sigma^2) - \frac{1}{2}E_{q_v}\Bigg\{\Bigg[\int_0^\infty f(\psi) - \int_0^{d^2} f(\psi)\Bigg]\Bigg[\Big((\sigma\sqrt{\psi}+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2$$

$$/(n+\delta)\Big)/(n+\gamma)\Bigg] \times \Bigg[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Big((\sigma\sqrt{\psi}+\sigma\sqrt{2}\lambda_1)^2$$

$$-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Bigg]\ d\psi\ \Bigg\}/\sigma^4 \qquad . \qquad (7.A.10)$$

Note that in the above expression, ω is denoted as $\sqrt{\psi}$, rather than $\psi^{1/2}$ to indicate that the sign of ω is positive (since the range of ω is restricted to $[d, \infty)$ in 7.A.9).

From (7.A.10), for any given q_v such that $cq_v/\sqrt{v}+\sqrt{2}\lambda_1 \le 0$,

$$\frac{\partial\rho(\hat{\sigma}^2,\ \sigma^2)}{\partial c} = \frac{1}{2}\left\{\frac{\partial}{\partial c}\int_0^{d^2} f(\psi)\left[\left((\sigma\sqrt{\psi}+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right] \times \right.$$
$$\left[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \left((\sigma\sqrt{\psi}+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2\right.\right.$$
$$\left.\left.\left./(n+\delta)\right)/(n+\gamma)\right]\ d\psi\ \right\}/\sigma^4 \tag{7.A.11}$$

$$= \frac{1}{2}\left\{\frac{\partial d^2}{\partial c}\ f(d^2)\left[\left((\sigma d+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right] \times \right.$$
$$\left[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \left((\sigma d+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2\right.\right.$$
$$\left.\left.\left./(n+\delta)\right)/(n+\gamma)\right]\right\}/\sigma^4 \quad . \tag{7.A.12}$$

Using the fact that $d = -cq_v/\sqrt{v} - \sqrt{2}\lambda_1$, which is non-negative, we can show that, for any given q_v,

$$\frac{\partial\rho(\hat{\sigma}^2,\ \sigma^2)}{\partial c} = \left\{(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)(q_v/\sqrt{v})f(d^2)\left[\left((\sigma(-cq_v/\sqrt{v} - \sqrt{2}\lambda_1)\right.\right.\right.$$
$$\left.\left.+\sigma\sqrt{2}\lambda_1)^2 \quad -\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right] \times \left[2\sigma^2(q_v^2-(n+\delta))\right.$$
$$/(n+\delta) + \left((\sigma(-cq_v/\sqrt{v} - \sqrt{2}\lambda_1)+\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2\right.$$
$$\left.\left.\left./(n+\delta)\right)/(n+\gamma)\right]\right\}/\sigma^4$$

$$= \left\{(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)(q_v/\sqrt{v})f(d^2)\left[\left(c^2\sigma^2q_v^2/v \quad -\sigma^2(\gamma-\delta)q_v^2\right.\right.\right.$$
$$\left.\left./(n+\delta)\right)/(n+\gamma)\right] \times \left[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \left(c^2\sigma^2q_v^2/v -\right.\right.$$
$$\left.\left.\left.\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right]\right\}/\sigma^4$$

$$= \Bigg\{(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)(q_v/\sqrt{v})f(d^2)\Bigg[\Bigg(\sigma^2 q_v^2(c^2/v - (\gamma-\delta)/(n+\delta))\Bigg)/(n+\gamma)\Bigg] \times \Bigg[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Bigg(\sigma^2 q_v^2(c^2/v - (\gamma-\delta)/(n+\delta)\Bigg)/(n+\gamma)\Bigg]\Bigg\}/\sigma^4 \quad . \quad (7.A.13)$$

Sufficient conditions for (7.A.13) to be zero are $c = -\infty$, which implies $f(d^2) = 0$, or $c^2/v - (\gamma-\delta)/(n+\delta) = 0$, *i.e.* $c = -\sqrt{v(\gamma-\delta)/(n+\delta)}$. This condition reduces to $c = 0$ for ML, $c = -1$ for LS and $c = -\sqrt{v/(v+2)}$ for MM.

Next, consider the case where $cq_v/\sqrt{v} + \sqrt{2}\lambda_1 > 0$. For the purposes of this analysis, we re-define d as $|cq_v/\sqrt{v} + \sqrt{2}\lambda_1| = cq_v/\sqrt{v} + \sqrt{2}\lambda_1 > 0$. Now, when $cq_v/\sqrt{v} + \sqrt{2}\lambda_1 > 0$,

$$\rho(\hat{\sigma}^2, \sigma^2) = \rho(\sigma^{**2},\sigma^2) - E\Bigg\{I_{(-\infty,d)}(\omega)\Bigg[\Bigg((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Bigg)/(n+\gamma)\Bigg] \times \Bigg[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \Bigg((\sigma\omega-\sigma\sqrt{2}\lambda_1)^2 -\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Bigg)/(n+\gamma)\Bigg]\Bigg\}/\sigma^4 \quad . \quad (7.A.14)$$

Proposition 2

For any given q_v, if $d \geq 0$, then

$$E\Big[I_{(-\infty,d)}(\omega)\omega^j\Big] = \begin{cases} g_j - E\Big[I_{[d,\infty}(\omega)\omega^j\Big] & \text{if j is even or zero} \\ - E\Big[I_{[d,\infty)}(\omega)\omega\Big] & \text{if j is odd} \end{cases} \quad , \quad (7.A.15)$$

where $g_j = 2^{j/2}\Gamma((j+1)/2)\Gamma(\frac{1}{2})$.

Proof :

$$E\Big[I_{(-\infty,d)}(\omega)\omega^j\Big] = \int_{-\infty}^{d} \frac{\omega^j e^{-\omega^2/2}}{\sqrt{2\pi}}\, d\omega$$

$$= \int_{-\infty}^{-d} \frac{\omega^{j} e^{-\omega^2/2}}{\sqrt{2\pi}} \, d\omega + \int_{-d}^{d} \frac{\omega^{j} e^{-\omega^2/2}}{\sqrt{2\pi}} \, d\omega \quad .$$

It is easy to see that $\frac{\omega^{j} e^{-\omega^2/2}}{\sqrt{2\pi}}$ is an odd function if j is odd, an even function otherwise. Using the properties of odd and even functions,

$$\int_{-\infty}^{-d} f(x)\,dx = \begin{cases} \displaystyle\int_{d}^{\infty} f(x)\,dx & \text{if f is an even function} \\ -\displaystyle\int_{d}^{\infty} f(x)\,dx & \text{if f is an odd function} \end{cases} \quad . \tag{7.A.16}$$

and

$$\int_{-d}^{d} f(x)\,dx = \begin{cases} 2\displaystyle\int_{0}^{d} f(x)\,dx & \text{if f is an even function} \\ 0 & \text{if f is an odd function} \end{cases} \quad . \tag{7.A.17}$$

Applying these results, when j is odd, we have

$$E\left[I_{(-\infty,d)}(\omega)\omega^{j}\right] = - E\left[I_{(d,\infty)}(\omega)\omega^{j}\right] \quad . \tag{7.A.18}$$

Alternatively, when j is even or zero,

$$\begin{aligned} E\left[I_{(-\infty,d)}(\omega)\omega^{j}\right] &= E\left[I_{(d,\infty)}(\omega)\omega^{j}\right] + 2E\left[I_{(0,d)}(\omega)\omega^{j}\right] \\ &= E\left[I_{(d,\infty)}(\omega)\omega^{j}\right] + 2\left\{E\left[I_{(0,\infty)}(\omega)\omega^{j}\right] - E\left[I_{(d,\infty)}(\omega)\omega^{j}\right]\right\} \\ &= 2E\left[I_{(0,\infty)}(\omega)\omega^{j}\right] - E\left[I_{(d,\infty)}(\omega)\omega^{j}\right] \end{aligned}$$

$$= g_j - E\left[I_{(d,\infty)}(\omega)\omega^j\right] \qquad (7.A.19)$$

Therefore,

$$E\left[I_{(-\infty,d)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^2\right] = \sigma^2 g_2 + \sigma^2 2\lambda_1^2 g_0 - E\left[I_{(d,\infty)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^2\right] \qquad (7.A.20)$$

$$E\left[I_{(-\infty,d)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^4\right] = \sigma^4 g_4 + \sigma^4 4\lambda_1^4 g_0 + 12\sigma^2\lambda_1^2 g_2 - E\left[I_{(d,\infty)}(\omega)(\sigma\omega - \sigma\sqrt{2}\lambda_1)^4\right] \qquad (7.A.21)$$

Using these results and the fact that $E\left[I_{(d,\infty)}(\omega)\omega^j\right] = \frac{1}{2}\left[\int_0^\infty f(\psi)\ \psi^{j/2}\ d\psi - \int_0^{d^2} f(\psi)\ \psi^{j/2}\ d\psi\right]$, when $cq_v/\sqrt{v}+\sqrt{2}\lambda_1 > 0$, (7.A.14) can be re-written as

$$\begin{aligned}\rho(\hat{\sigma}^2, \sigma^2) = \rho(\sigma^{**2}, \sigma^2) - E_{q_v}\Big[&\sigma^4 g_4 - 12\sigma^2\lambda_1^2 g_2 - 4\lambda_1^4 g_0 - \sigma^4(\gamma-\delta)^2 q_v^4/(n+\delta)^2 + \\ &2\sigma^2(\gamma-\delta)q_v^2(\sigma^2 g_2 + 2\lambda_1^2 g_0)/(n+\delta)\Big]/(n+\gamma)^2 - 2\sigma^2(q_v^2 - (n+\delta)) \\ &\left[\sigma^2 g_2 + 2\lambda_1^2 g_0 - \sigma^2(\gamma-\delta)q_v^2 g_0/(n+\delta)\right]/((n+\delta)(n+\gamma)) \\ &+ \frac{1}{2} E_{q_v}\left[\int_0^\infty f(\psi)\ d\psi - \int_0^{d^2} f(\psi)\ d\psi\right]\left[\left[\left((\sigma\sqrt{\psi}-\sigma\sqrt{2}\lambda_1)^2\right.\right.\right. \\ &\left.\left. -\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right] \times \left[2\sigma^2(q_v^2-(n+\delta))/(n+\delta) + \right. \\ &\left.\left.\left((\sigma\sqrt{\psi}-\sigma\sqrt{2}\lambda_1)^2 - \sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right]\right\} f(\psi)\ /\sigma^4 .\end{aligned}$$

(7.A.22)

Hence, for any q_v such that $cq_v/\sqrt{v} + \sqrt{2}\lambda_1 > 0$,

$$\frac{\partial\rho(\hat{\sigma}^2, \sigma^2)}{\partial c} = \frac{1}{2}\left\{\frac{\partial}{\partial c}\int_0^{d^2} f(\psi)\left[\left((\sigma\sqrt{\psi}-\sigma\sqrt{2}\lambda_1)^2 - \sigma^2(\gamma-\delta)q_v^2/(n+\delta)\right)/(n+\gamma)\right]\right.$$

$$\times\Big[2\sigma^2(q^2-(n+\delta))/(n+\delta) + \Big((\sigma\sqrt{\psi}-\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2$$

$$/(n+\delta)\Big)/(n+\gamma)\Big]\ d\psi\ \Big\}/\sigma^4$$

$$= \Big\{(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)(q_v/\sqrt{v})f(d^2)\Big[\Big((\sigma(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)$$

$$-\sigma\sqrt{2}\lambda_1)^2 \quad -\sigma^2(\gamma-\delta)q_v^2/(n+\delta)\Big)/(n+\gamma)\Big] \times \Big[2\sigma^2(q^2-(n+\delta))$$

$$/(n+\delta) + \Big((\sigma(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)-\sigma\sqrt{2}\lambda_1)^2-\sigma^2(\gamma-\delta)q_v^2$$

$$/(n+\delta)\Big)/(n+\gamma)\Big]\Big\}/\sigma^4$$

$$= \Big\{(cq_v/\sqrt{v} + \sqrt{2}\lambda_1)(q_v/\sqrt{v})f(d^2)\Big[\Big(\sigma^2q_v^2(c^2/v - (\gamma-\delta)$$

$$/(n+\delta))\Big)/(n+\gamma)\Big] \times \Big[2\sigma^2(q^2-(n+\delta))/(n+\delta) + \Big(\sigma^2q_v^2(c^2/v -$$

$$(\gamma-\delta)/(n+\delta))\Big)/(n+\gamma)\Big]\Big\}/\sigma^4 \quad . \qquad (7.A.23)$$

Consistent with the case where $cq_v+\sqrt{2}\lambda_1 \le 0$, a sufficient condition for (7.B.23) to be zero is $c^2/v - (\gamma-\delta)/(n+\delta) = 0$, *i.e.* $c = -\sqrt{v(\gamma-\delta)/(n+\delta)}$ or $c = -\infty$. Note that these results do not depend on λ_2, and hence they also apply to the case where the model is well specified.

APPENDIX 7B

Figure 7.1

Relative risk functions of $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML}, σ^{**2}_{ML} and $\hat{\sigma}^2_{ML}$ for n = 20 k = 5 and λ_2 = 2

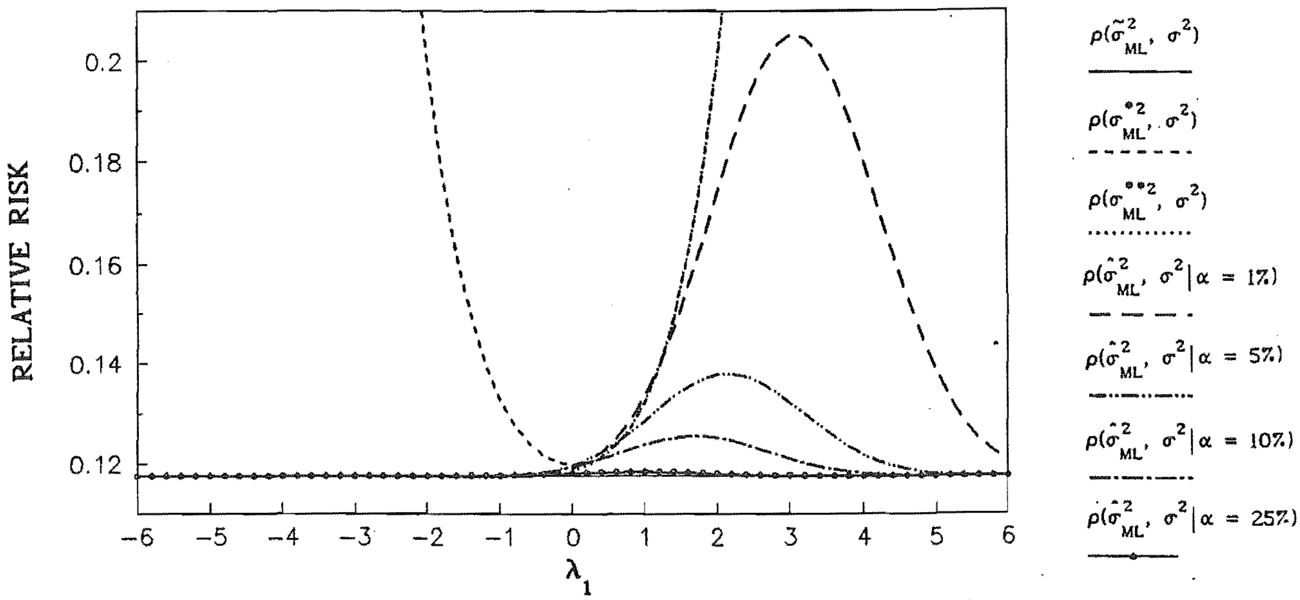


Figure 7.2

Relative risk functions of $\tilde{\sigma}^2_{ML}$, σ^{*2}_{ML}, σ^{**2}_{ML} and $\hat{\sigma}^2_{ML}$ for n = 20 k = 5 and λ_2 = 10

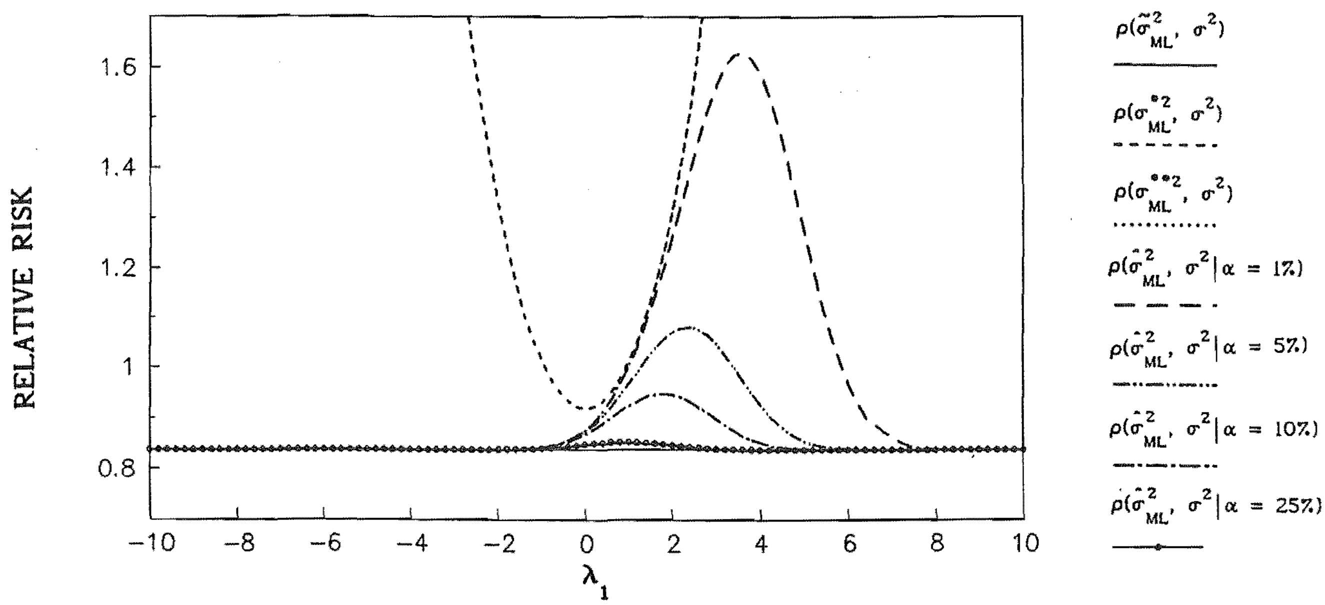


Figure 7.3

Relative risk functions of $\tilde{\sigma}^2_{LS}$, σ^{*2}_{LS}, σ^{**2} and $\hat{\sigma}^2_{LS}$ for n = 20 k = 5 and λ_2 = 2

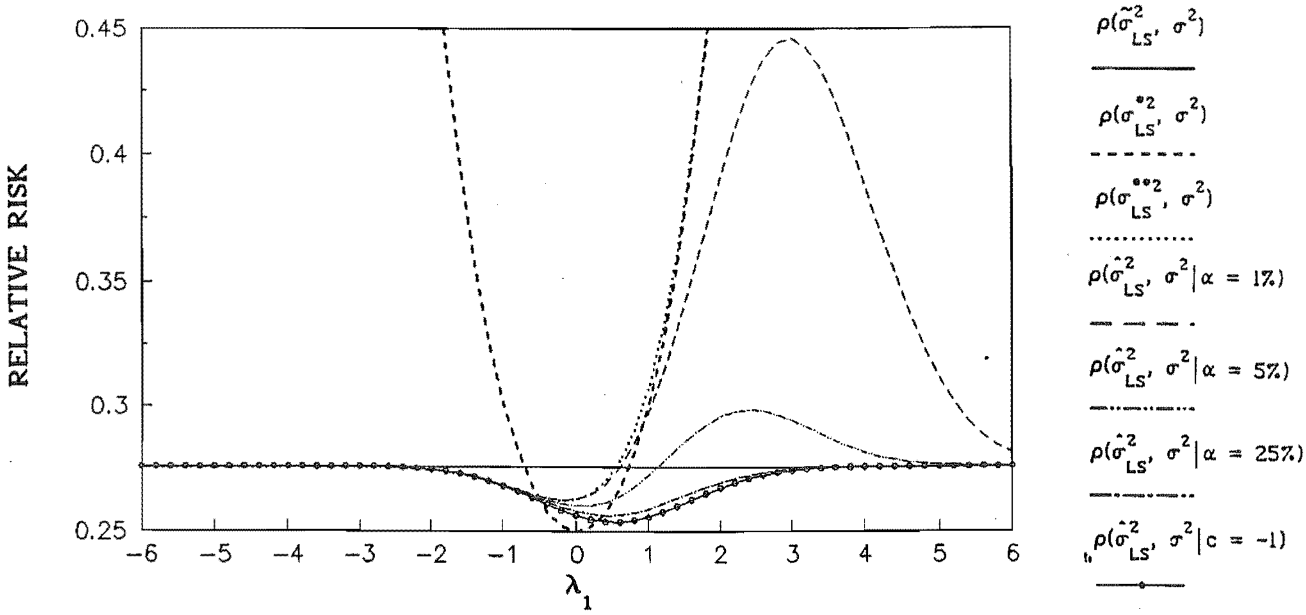


Figure 7.4

Relative risk functions of $\tilde{\sigma}^2_{LS}$, σ^{*2}_{LS}, σ^{**2}_{LS} and $\hat{\sigma}^2_{LS}$ for n = 20 k = 5 and λ_2 = 10

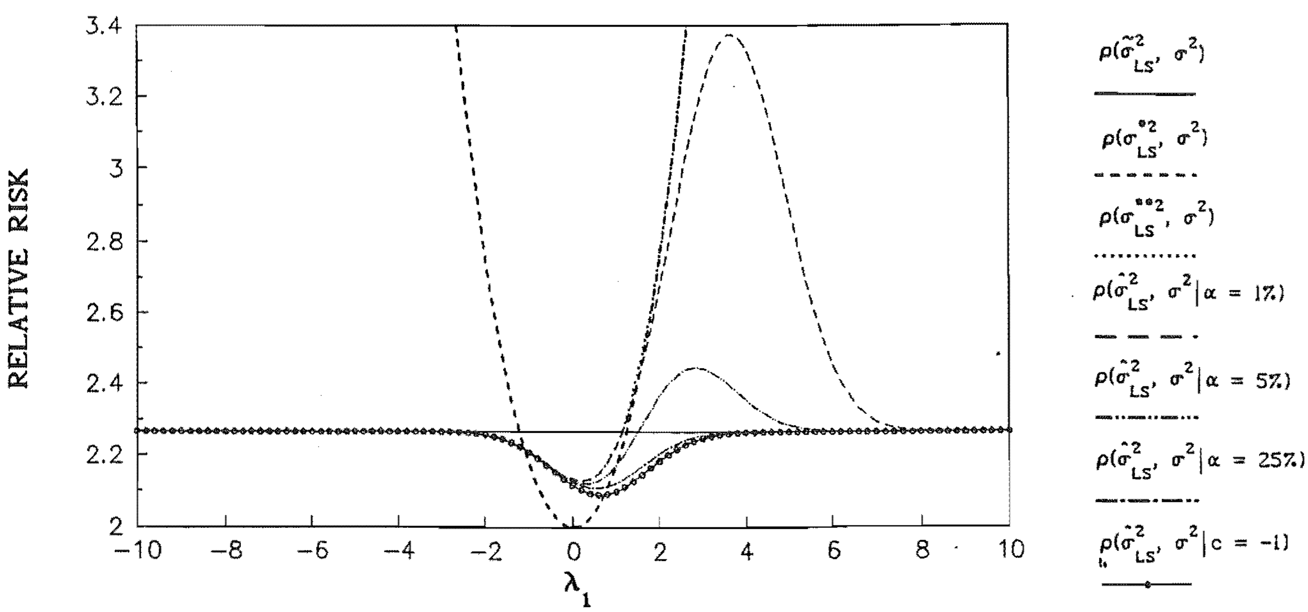


Figure 7.5

Relative risk functions of $\tilde{\sigma}^2_{MM}$, σ^{*2}_{MM}, σ^{**2}_{MM} and $\hat{\sigma}^2_{MM}$ for n = 20 k = 5 and λ_2 = 2

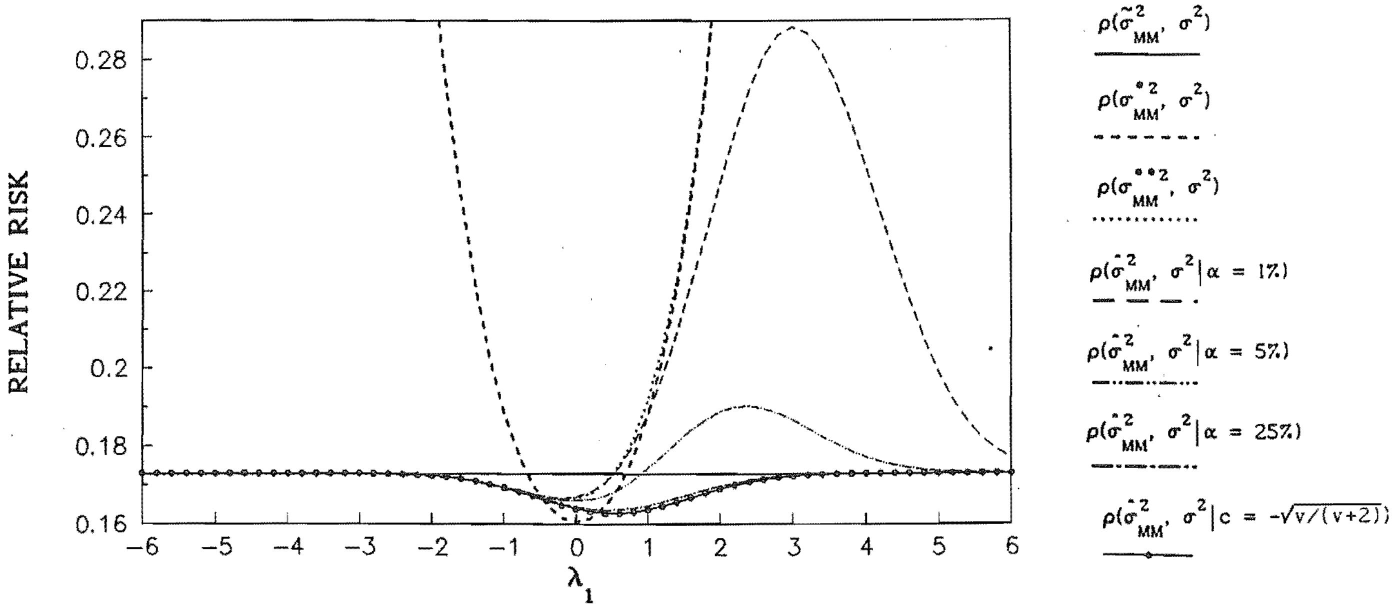
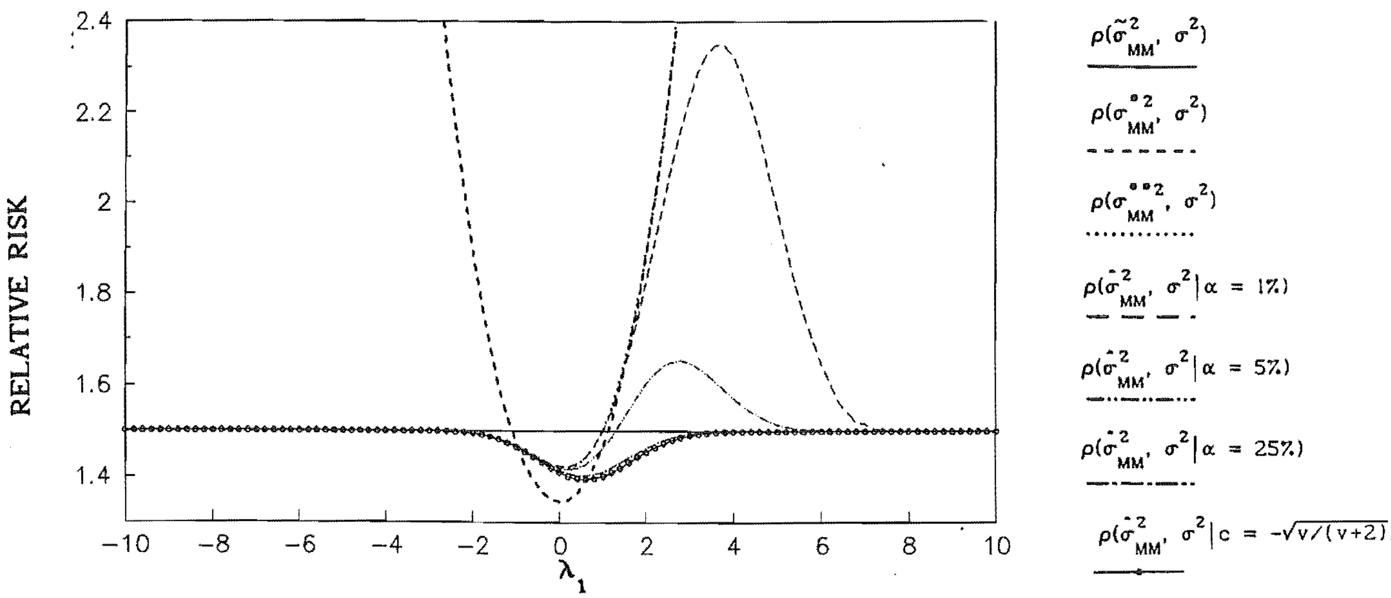


Figure 7.6

Relative risk functions of $\tilde{\sigma}^2_{MM}$, σ^{*2}_{MM}, σ^{**2}_{MM} and $\hat{\sigma}^2_{MM}$ for n = 20 k = 5 and λ_2 = 10



APPENDIX 7C

The properties of the inequality restricted and pre-test estimators when irrelevant regressors are included in the model

Consider the linear regression model

$$y = X_1\beta_1 + \varepsilon \quad , \qquad \varepsilon \sim N(0, \sigma^2 I) \qquad (7.C.1)$$

where y and ε are $n \times 1$, β_1 is $k_1 \times 1$, X_1 is $n \times k_1$, non-stochastic and of rank k.

Suppose the fitted model is

$$y = X_1\beta_1 + X_2\beta_2 + \mu = X\beta + \mu \quad , \qquad (7.C.2)$$

where X_2 is $n \times k_2$ and of rank k_2, β_2 is $k_2 \times 1$ and $k = k_1 + k_2$.

The prior information is given by

$$C'\beta \geq r \quad . \qquad (7.C.3)$$

As in Chapter 4 and 6, we define an orthonormal matrix Q such that $QS^{-1/2}C(C'S^{-1}C)^{-1}C'S^{-1/2}Q' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, where $S = X'X$. Letting $h' = C'S^{-1/2}Q'$, we can show that $h' = \begin{bmatrix} h_1, & 0' \end{bmatrix}$.

Using these results, we may reparameterize (7.C.2) and (7.C.3) into the orthonormal model

$$y = XS^{-1/2}Q'QS^{1/2}\beta + \mu = H\theta + \mu \ , \qquad (7.C.4)$$

where $\theta = QS^{1/2}\beta$ and $H = XS^{-1/2}Q'$.

Noting that $\beta = (\beta_1, \beta_2)'$ and that $\beta_2 = 0$, we see that

$$\theta = QS^{1/2}\begin{bmatrix} \beta_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \zeta \\ \pi \end{bmatrix}. \qquad (7.C.5)$$

Accordingly, $\theta_1 = \zeta_1$, where θ_1 and ζ_1 are the first elements in θ and ζ respectively. Note that if (7.C.2) is the true model, then $\beta_2 \neq 0$, $\theta \neq \begin{bmatrix} \zeta \\ \pi \end{bmatrix}$ and hence $\theta_1 \neq \zeta_1$.

Using the mechanism presented in Chapter 4, we can reparameterize the inequality restriction as

$$\theta_1 + \tau = r_0 \quad , \qquad (7.C.6)$$

where r_0 is defined as in Chapter 4, and $\tau \le 0$ if the constraint is correct. When the model is overfitted, $\theta_1 = \zeta_1$ and thus $\tau = \tau_1 = r_0 - \zeta_1$. If (7.C.2) is the correct specification, then $\tau \ne \tau_1$.

Analgously to the results presented in Chapters 4, 6 and 7, the inequality restricted estimator of β and σ^2 can be shown to be

$$\beta^{**} = \tilde{\beta} - S^{-1/2}Q'\begin{bmatrix} I_{(-\infty,\tau_1/\sigma)}(u_1)(\sigma u_1 - \tau_1) \\ 0 \end{bmatrix} \tag{7.C.7}$$

and

$$\sigma^{**2} = \tilde{\sigma}^2 + I_{(-\infty,\tau_1/\sigma)}(u_1)\left[\left((\sigma u_1 - \tau_1)^2 - \tilde{\sigma}^2(\gamma - \delta)\right)/(n+\gamma)\right] \tag{7.C.8}$$

respectively, where $u_1 = (\tilde{\theta}_1 - \theta_1)/\sigma$ and δ and γ are defined as in Chapters 6 and 7.

$$\begin{aligned} \text{Now } E(\tilde{\theta}) &= QS^{1/2}E(\tilde{\beta}) \\ &= QS^{1/2}E\left[S^{-1}X'(X_1\beta_1 + \varepsilon)\right] \\ &= QS^{1/2}S^{-1}X'(X_1\beta_1 + E(\varepsilon)) \\ &= QS^{1/2}\begin{bmatrix} I_{(k_1\times k_1)} \\ 0_{(k_2\times k_1)} \end{bmatrix}\beta_1 \\ &= \theta \end{aligned} \tag{7.C.9}$$

and

$$E\left[(\tilde{\theta} - E(\tilde{\theta}))(\tilde{\theta} - E(\tilde{\theta}))'\right] = \sigma^2 I.$$

Also, given that ε is normally distributed, $\tilde{\theta}$ is also normally distributed. Accordingly, $u_1 \sim N(0, 1)$ as in the case of a properly specified model.

Also, as $\tilde{\theta} \sim N(\theta, \sigma^2 I)$, it can be shown easily that $(n+\delta)\tilde{\sigma}^2/\sigma^2 \sim \chi^2_{n-k}$, as in the case when (7.C.2) is a correctly specified model.

Given these results, it is apparent that the properties of β^{**} and σ^{**2} are the same as when (7.C.2) is correctly specified, but with τ_1 replacing τ everywhere.

Furthermore, when the model is overfitted, the test statistic for testing (7.C.3) is given as

$$t = \sqrt{n+k}(\tilde{\theta}_1 - r_0)/(\tilde{\sigma}\sqrt{n+\delta}) \quad .$$

As $\tilde{\theta}_1 \sim N(\theta_1, \sigma^2)$, $(\tilde{\theta}_1 - r_0)/\sigma$ has a standard normal distribution when $\theta_1 = r_0$ as when the model is well specified. $\tilde{\sigma}^2(n+\delta)/\sigma^2$ has a chi-square distribution with n - k degrees of freedom and $\tilde{\theta}$ and $\tilde{\sigma}^2$ are independently distributed. Therefore, when $\theta_1 = r_1$, t has a Student's *t* distribution with n - k degrees of freedom as in the case of a well specified model.

Given these results, it seems reasonable to conjecture that the properties of the inequality pre-test estimators in an overfitted model are essentially the same as in a correctly specified model, except that the surplus variable τ is scaled to τ_1 when the model is overfitted. The derivations of risk functions under the assumption of an overfitted model therefore appear unncessary. The conclusion given here is in accord with the findings of Giles (1986) when the constraint holds as an equality.

CHAPTER EIGHT

THE OPTIMAL CRITICAL VALUE FOR A PRE-TEST OF AN INEQUALITY RESTRICTION WHEN ESTIMATING THE SCALE PARAMETER

8.1 INTRODUCTION

In Chapter 5 of this thesis, using the criteria of minimum average relative risk and mini-max regret, we derived and presented optimal critical values for a pre-test of $C'\beta \geq r$ when estimating $E(y)$. We found that the mini-max regret critical value depends on the model's degrees of freedom when the model is underfitted, but it is approximately constant when the model is well specified. If the alternative minimum average relative risk criterion is adopted, we showed that the inequality pre-test estimator with the optimal pre-test size is in fact the unrestricted estimator when only one inequality restriction is being tested.

In this chapter, we extend the analysis of Chapter 5 to the case in which the researcher's focus is on the estimation of σ^2. Optimal critical values for the pre-test are derived and tabulated according to the mini-max regret and minimum average relative risk criteria. As in the case of estimating the prediction vector, this is considered within the context of an underfitted model, of which the standard linear model is a special case. We note from our results of the previous chapter that when one is estimating the error variance, underfitting the model can give rise to strictly dominating estimators within the class under consideration in this thesis. In these cases, the choice of optimal critical value is obvious. Accordingly, our discussion will concentrate on the cases in which there exists no strictly dominating estimator. As in Chapter 5, we will use c^{MX} and c^{A} to denote the optimal critical values derived using the mini-max regret and minimum average relative risk criteria respectively. The subscripts ML, LS and MM are used to characterize the three component estimators of σ^2 as in Chapter 6 and 7.

8.2 THE CHOICE OF AN OPTIMAL CRITICAL VALUE ACCORDING TO THE CRITERION OF MINIMUM AVERAGE RELATIVE RISK

By analogy with the definition given in Chapter 5, for any given λ_1, the relative risk of the inequality pre-test estimator $\hat{\sigma}^2$ is defined as

$$R(\hat{\sigma}^2) = \rho(\hat{\sigma}^2, \sigma^2) - \min_c \rho(\hat{\sigma}^2, \sigma^2) \quad , \tag{8.1}$$

where $\min_c \rho(\hat{\sigma}^2, \sigma^2)$ denotes the minimum of $\rho(\hat{\sigma}^2, \sigma^2)$ over all critical values, and is therefore the minimum risk boundary of $\hat{\sigma}^2$. The criterion of minimizing the average relative risk aims to choose a critical value that minimizes (8.1) over the entire range of λ_1 in some average sense. That is, we want to minimize

$$A(c) = \int_{-\infty}^{\infty} \left[\rho(\hat{\sigma}^2, \sigma^2) - \min_c \rho(\hat{\sigma}^2, \sigma^2)\right] d\lambda_1 \geq 0 . \tag{8.2}$$

Now, from the results reported in Chapters 6 and 7, for a relatively small λ_2 (say, ≤ 5), $\min_c \rho(\hat{\sigma}^2_{LS}, \sigma^2) = \min \left[\rho(\hat{\sigma}^2_{LS}, \sigma^2 | c = -\infty), \rho(\hat{\sigma}^2_{LS}, \sigma^2 | c = -1)\right]$ and $\min_c \rho(\hat{\sigma}^2_{MM}, \sigma^2) = \min \left[\rho(\hat{\sigma}^2_{MM}, \sigma^2 | c = -\infty), \rho(\hat{\sigma}^2_{MM}, \sigma^2 | c = -\sqrt{v/(v+2)})\right]$ for any given λ_1. Clearly, a change in the size of the pre-test would not alter the minimum risk boundary of $\hat{\sigma}^2$. In otherwords, $\min_c \rho(\hat{\sigma}^2, \sigma^2)$ is invariant to c. Hence, $\frac{\partial}{\partial c} \min_c \rho(\hat{\sigma}^2, \sigma^2) = 0$, which implies $\frac{\partial}{\partial c} \int \min_c \rho(\hat{\sigma}^2, \sigma^2)\, d\lambda_1 = 0$. Therefore,

$$\frac{\partial A(c)}{\partial c} = \frac{\partial}{\partial c} \int_{-\infty}^{\infty} \rho(\hat{\sigma}^2, \sigma^2)\, d\lambda_1 \quad . \tag{8.3}$$

Since $\frac{\partial}{\partial c}\int_{-\infty}^{\infty} \rho(\hat{\sigma}^2, \sigma^2)d\lambda_1$ can be written[1] as $\int_{-\infty}^{\infty} \frac{\partial}{\partial c} \rho(\hat{\sigma}^2,\sigma^2)\ d\lambda_1$, a sufficient (but not necessary) condition for (8.3) to be zero is that $\frac{\partial}{\partial c} \rho(\hat{\sigma}^2,\sigma^2) = 0$.

Recall from Appendix 7A that $\frac{\partial}{\partial c} \rho(\hat{\sigma}^2,\sigma^2) = 0$ when $c = 0, -\infty$ in the case of ML, when $c = -1, -\infty$ in the case of LS, and when $c = -\sqrt{v/(v+2)}, -\infty$ in the case of MM. Since $A(-\infty)$ does not converge, it is clear that $c = -\infty$ does not yield a minimum. Using a grid search procedure similar to that described in Chapter 5, we have computed the values of c for which A(c) is minimized for v = 2, 15, 25, 50 and λ_2 = 0, 2, 10, 25 for the case of LS and MM, and for n = 20, 30, 40, 50, 80, k = 2, 5, 10, 15 ... (n-5) and λ_2 = 0, 2, 10, 25, 50 for the case of[2] ML.

Our numerical results suggest that at least for the cases that we have considered, regardless of the chosen values of v and λ_2, A(c) reaches a minimum at $c = -1$ for the LS case and at $c = -\sqrt{v/(v+2)}$ for the MM case. In a sense, this result is hardly surprising given that the pre-test estimator corresponding to $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM) is the minimum risk estimator over much of the parameter space, and that the degree to which $\rho(\hat{\sigma}^2_{LS}, \sigma^2|c = -1)$ and $\rho(\hat{\sigma}^2_{LS}, \sigma^2|c = -\sqrt{v/(v+2)})$ are dominated by the inequality restricted estimators of their respective families over the rest of the parameter space is typically very slight. However, unlike most previous results, the optimal critical values presented here are invariant to changes in other parameters of the model. Tables 8.1 and 8.2 in Appendix 8A (see pp.

---

[1] See, for example, Bartle (1966, p. 46) for a proof.

[2] We use the trapezoidal rule to approximate A(c) for various values of c. We divided the interval of λ_1 into segments of 0.02. The upper and lower limits of λ_1 were selected such that the values of $\rho(\hat{\sigma}^2, \sigma^2)$ corresponding to both limits converge to the risk of the unrestricted estimator. These calculations were preformed using double precision FORTRAN code executed on a VAX 6340 computer.

201-203) illustrate some of the results obtained from our grid search. Although the results presented here are not based on analytical methods, and are confined only to certain values of v and λ_2, given the properties of the pre-test risk function, it is likely that our findings will hold for all values of v and λ_2.

When the component estimator is ML, the optimal critical value varies with the number of regression coefficients and observations, and with the model specification error. For a given n, $|c^A_{ML}|$ increases with the number of regression coefficients, but decreases with the model specification error. For any given k and λ_2, c^A_{ML} is fairly constant as n varies. Given n and k, $|c^{ML}| \to 0$ as $\lambda_2 \to \infty$, reflecting the fact that $\tilde{\sigma}^2_{ML}$ strictly dominates when the prior information is significantly incorrect. Table 8.3 illustrates some of these results (see p. 205).[3]

8.3 THE CHOICE OF AN OPTIMAL CRITICAL VALUE ACCORDING TO THE CRITERION OF MINI-MAX REGRET

The other optimality criterion that we consider in this chapter is that of mini-max regret. Analgous to the definition adopted in Chapter 5, for any given λ_1 and c, we define the regret function of the risk of the inequality pre-test estimator as

$$\mathrm{REG}(\lambda_1,\ c) = \rho(\hat{\sigma}^2,\ \sigma^2) - \min_c \rho(\hat{\sigma}^2,\ \sigma^2) \qquad . \tag{8.4}$$

As noted above,

$$\min_c \rho(\hat{\sigma}^2,\ \sigma^2) = \begin{cases} \rho(\hat{\sigma}^2,\ \sigma^2|c = -\infty) & \text{if } \lambda_1 \le \lambda_1^* \\ \\ \rho(\hat{\sigma}^2,\ \sigma^2|c = c^*) & \text{if } \lambda_1 > \lambda_1^* \end{cases} \quad , \tag{8.5}$$

---

[3] In Table 8.3 we only show the optimal critical value and the corresponding average relative risk in order to avoid formidable output. Complete results of our grid search are available on request.

where $c^* = 0$ (for ML), -1 (for LS) and $-\sqrt{v/(v+2)}$ (for MM). We denote d^L as the maximum of $\mathrm{Reg}(\lambda_1, c)$ in the region $\lambda_1 \le \lambda_1^*$, and d^U as the maximum of the regret function in the region $\lambda_1 > \lambda_1^*$, where λ_1^* is the value of $\lambda_1 > 0$ at which $\rho(\hat{\sigma}^2, \sigma^2 | c = -\infty)$ intersects $\rho(\hat{\sigma}^2, \sigma^2 | c = c^*)$.

When the component estimator is ML, the minimum risk boundary of the inequality pre-test estimator is given by the unrestricted and the inequality restricted estimator; *i.e.*, the inequality pre-test estimators that correspond to the two extreme critical values. Furthermore, increasing $|c|$ decreases d^L, but increases[4] d^U. Because of this monotonicity property, the mini-max regret procedure is to find the critical value c^{MX} such that $d^L = d^U$. That is, both regrets are simultaneously minimized. It is readily shown that c^{MX} is unique.

We performed numerical computations to calculate c^{MX}_{ML} for n = 20, 30, 40, 50, 80, k = 2, 5, 10, 15,...(n-5), $\lambda_2 = 0, 2, 10, 25, 50$. As in Chapter 5, Brent's (1974) algorithm was used to search for the value of λ_1^*. The Golden Section Search Routine given in Press *et al.* (1986) was used to compute the mini-max regret critical values. These were incoropranted into a double precision Fortran programme executed on a VAX 6340 computer. Table 8.4 illustrates the results (see p. 206). It is shown that for any given n, $|c^{MX}_{ML}|$ increases as k increases, *i.e.*, c^{MX}_{ML} is not invariant to the model's degrees of freedom. This differs from the results that we observed when estimating E(y), but it is consistent with the results of Giles and Lieberman (1991) who consider the case where the linear restriction is in the form of a strict

---

[4] The exception to this rule is when changes in c lie within the ranges such that the corresponding pre-test estimators strictly dominate the unrestricted estimator, in which case d^L decreases as $|c|$ decreases, but d^U remains unchanged. However, given that $\hat{\sigma}^2$ is never the minimum risk estimator, and that any further increase in $|c|$ will ultimately increase d^U, the critical value that equals d^U and d^L is still the mini-max regret critical value.

equality. For any given k, c_{ML}^{MX} is roughly constant as n varies. Other things being equal, c_{ML}^{MX} approaches zero as λ_2 approaches infinity. These findings are qualitatively the same as those reported in the previous section when the optimality criterion adopted is that of minimum average relative risk. Quantitatively though, for any given n, k and λ_2, $|c_{ML}^{MX}| \geq |c_{ML}^{A}|$. The use of the mini-max regret critical value would therefore result in a more frequent acceptance of the null, other things being equal. Some representative diagrams depicting the risk functions of $\hat{\sigma}^2_{ML}$ corresponding to c_{ML}^{MX} and c_{ML}^{A} are given in Appendix 8B (see pp. 211-213).

When the component estimator is LS or MM, for a relatively small λ_2, the minimum risk of $\hat{\sigma}^2$ is given by $\hat{\sigma}^2|c = -\infty$, or equivalently, σ^{**2}, in the region $\lambda_1 \leq \lambda_1^*$. Consequently, a decrease in c always brings the inequality pre-test estimator closer to the inequality restricted estimator and *vice versa*. Hence d^L decreases monotonically with c, as in the case when the ML component estimator is used. By contrast, when $\lambda_1 > \lambda_1^*$, minimum risk is achieved by the pre-test estimator with $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM). As both -1 and $-\sqrt{v/(v+2)}$ lie between the extremes of 0 and $-\infty$, the direction of change in risk which results from increasing/decreasing c cannot be determined. Consequently, d^U does not generally change monotonically with c. In other words, when estimating σ^2 based on the MM or LS principles, decreasing (increasing) $|c|$ increases (decreases) d^L, but does *not* necessarily decrease (increase) d^U. Accordingly, the critical value that equalizes d^L and d^U is not necessarily mini-max regret. This constrasts with the case when one is estimating E(y), or estimating σ^2 using the ML component.

The procedure of seeking a mini-max regret critical value therefore is to search for a critical value that minimizes the maximum of the regret function over the entire range of λ_1. Now, let d^{*L} denote the maximum difference between $\rho(\hat{\sigma}^2, \sigma^2|c = c^*)$ and $\rho(\sigma^{**2}, \sigma^2)$ in the region $\lambda_1 \le \lambda_1^*$. d^{*L} is therefore the maximum regret of using $\hat{\sigma}^2|c = c^*$ rather than σ^{**2} when $\lambda_1 \le \lambda_1^*$. When $\lambda_1 > \lambda_1^*$, $\hat{\sigma}^2|c = c^*$ is the dominating estimator and accordingly $\mathrm{Reg}(\lambda_1, c^*) = 0$. Suppose there exists $\bar{c}$ such that $\bar{c} \ne c^*$, $\bar{d}^L < d^{*L}$ and $\bar{d}^U < d^{*L}$, where $\bar{d}^L$ and $\bar{d}^U$ are the maximum values of $\mathrm{Reg}(\lambda_1, \bar{c})$ in the region $\lambda_1 \le \lambda_1^*$ and $\lambda_1 > \lambda_1^*$ respectively. Then it is obvious that c^* does not minimize the maximum regret and hence cannot be the mini-max regret critical value. If $\bar{c}$ does not exist, then c^* is the mini-max regret critical value. Now suppose further that $\bar{c}$ is not unique. That is, there exists $\tilde{c}$ such that $\tilde{c} \ne \bar{c}$ and $\tilde{d}^L < d^{*L}$ and $\tilde{d}^U < d^{*L}$, where $\tilde{d}^L(\tilde{d}^U)$ is defined analgous to $\bar{d}^L(\bar{d}^U)$. Then under a mini-max regret criterion, $\bar{c}$ is preferred to $\tilde{c}$ (and is therefore the optimal critical value) if and only if $\max(\bar{d}^L, \bar{d}^U) < \max(\tilde{d}^L, \tilde{d}^u)$.

As it seems impossible to derive the mini-max regret critical value analytically, we resort to numerical computations. We consider $\lambda_2 = 0$, 2 and the same values of n, k as in the previous section. Given that the inequality pre-test estimators corresponding to $c = -1$ (for LS) or $c = -\sqrt{v/(v+2)}$ (for MM) strictly dominate all other inequality pre-test estimators of their respective families when λ_2 is large, as shown in Table 8.1 and 8.2, the mini-max regret critical value for cases of relatively large λ_2 is obvious and requires no discussion.

Our numerical results are illustrated in Tables 8.4 - 8.5 (pp. 206-207). They suggest that when the model is well specified and the LS or MM component estimators are used, at least for the parameter values that we have selected, the mini-max regret critical values are roughly constant for moderate to high degrees of freedom (approximately -1.14 for the case of LS and -1.76 for the case of MM). This is consistent with the results that one obtains when estimating E(y). When the possibility of omitted variables is allowed for, then the optimal critical value differs only marginally from $c = -1$ in the LS case and from $c = -\sqrt{v/(v+2)}$ in the MM case. Again, this result is not surprising, given that the pre-test estimators corresponding to these critical values are the minimum risk estimators over almost the entire λ_1 range. Our findings also suggest that when one is estimating using the LS or MM components, the optimal critical values resulting from the minimum average relative risk and mini-max regret criteria are roughly the same for relatively large λ_2. Some representative diagrams illustrating these results are given in Appendix 8B.

8.4 CONCLUSIONS

In this chapter, we have considered the question of the optimal choice of critical value for the pre-test when estimating the regression scale parameter according to the criteria of mini-max regret or minimum average relative risk. Perhaps the most striking feature of our results is that when estimating σ^2 using the LS or MM components, the optimal critical value according to the criterion of minimum average relative risk is always -1 for the LS case and $-\sqrt{v/(v+2)}$ for the MM case, regardless of the number of regressors and observations and the extent to which the model is mis-specified. This is not surprising, given that the pre-test estimator corresponding to $c = -1$ (for LS) and $c = -\sqrt{v/(v+2)}$ (for MM) is the minimum risk estimator over much of the parameter space, and the degree to which these pre-test estimators are dominated by the inequality restricted estimators of their respective family over the rest of the parameter space is typically very slight. When the model is correctly specified, this result is qualitatively consistent with that obtained by Toyoda and Wallace (1975) for a pre-test of variance homogeneity. Alternatively, when the criterion of mini-max regret is adopted, the optimal critical values are roughly constant with respect to the model's degrees of freedom.

By contrast, if the method of maximum likelihood is used, then the optimal critical value according to these criteria varies with the number of observations and regression coefficients in the model, as well as the degree of model mis-specification. In particular, the optimal critical value is not invariant to the model's degrees of freedom when the model is well specified. This contrasts with the results reported in Chapter 5 when one is estimating E(y), but concurs qualitatively with the results obtained by Giles and Lieberman (1991) when the pre-test in question is a strict equality constraint.

APPENDIX 8A

Table 8.1

Average relative risk A(c) of the inequality pre-test estimator (MM component)

		λ_2			
v	c	0	2	10	25
2	$-\infty$	∞	∞	∞	∞
	⋮	⋮	⋮	⋮	⋮
	-2.0	0.86984	11.65700	105.95121	169.15091
	-1.8	0.54644	7.69982	88.04181	166.89212
	-1.6	0.33493	4.61843	65.58911	159.68198
	-1.4	0.20699	2.44170	41.50487	139.91612
	-1.2	0.13696	1.06842	20.55817	98.64767
	-1.0	0.10400	0.32405	6.67456	42.83098
	-0.8	0.09273	0.02821	0.59491	4.47593
	-0.71*	0.09179	0.02672	0.00018	0
	-0.6	0.09270	0.03201	0.69022	5.46981
	-0.4	0.09752	0.22448	4.93870	94.01971
	-0.2	0.10388	0.52533	11.68413	124.83132
	0	0.11049	0.87537	19.59501	157.97205
15	$-\infty$	∞	∞	∞	∞
	⋮	⋮	⋮	⋮	⋮
	-2.0	0.02625	0.00997	0.87790	4.90101
	-1.8	0.01667	0.05932	0.53642	3.05642
	-1.6	0.01074	0.03158	0.29271	1.68759
	-1.4	0.00737	0.01387	0.13145	0.76401
	-1.2	0.00572	0.00401	0.03876	0.22694
	-1.0	0.00517	0.00020	0.00187	0.01143
	-0.94*	0.00514	1.81×10^{-6}	0	0
	-0.8	0.00525	0.00093	0.00931	0.05512
	-0.6	0.00565	0.00494	0.05082	0.30021
	-0.4	0.00624	0.01318	0.11708	0.69284
	-0.2	0.00680	0.01911	0.19946	1.18189
	0	0.00740	0.02759	0.28991	1.71882
25	$-\infty$	∞	∞	∞	∞
	⋮	⋮	⋮	⋮	⋮
	-2.0	0.00991	0.03177	0.22261	1.09712
	-1.8	0.00631	0.01878	0.13524	0.67172
	-1.6	0.00408	0.00987	0.07293	0.36435
	-1.4	0.00283	0.00422	0.03194	0.16028
	-1.2	0.00223	0.00113	0.00873	0.04402
	-1.0	0.00204	0.00003	0.00022	0.00102
	-0.96*	0.00204	4.18×10^{-7}	0	0
	-0.8	0.00210	0.00043	0.00032	0.01758

Table 8.1 (cont'd)

		λ_2			
v	c	0	2	10	25
25	-0.6	0.00227	0.00194	0.01584	0.08078
	-0.4	0.00251	0.00422	0.03491	0.17840
	-0.2	0.00275	0.00699	0.05835	0.29868
	0	0.00300	0.00999	0.08396	0.43020
50	$-\infty$	∞	∞	∞	∞
	⋮	⋮	⋮	⋮	⋮
	-2.0	0.00257	0.00715	0.03888	0.15506
	-1.8	0.00164	0.00420	0.02350	0.09432
	-1.6	0.00107	0.00219	0.01255	0.05063
	-1.4	0.00075	0.00091	0.00536	0.02175
	-1.2	0.00059	0.00027	0.00137	0.00557
	-1.0	0.00055	1.74×10^{-6}	1.01×10^{-5}	4.10×10^{-5}
	-0.98*	0.00055	6.65×10^{-8}	0	0
	-0.8	0.00057	0.00013	0.00079	0.00325
	-0.6	0.00062	0.00051	0.00324	0.01335
	-0.4	0.00069	0.00107	0.00693	0.02857
	-0.2	0.00076	0.00175	0.01414	0.04716
	0	0.00083	0.00249	0.01630	0.06743

* indicates the critical value that corresponds to $-\sqrt{v/(v+2)}$.

Table 8.2

Average relative risk A(c) of the inequality pre-test estimator (LS component)

		λ_2			
v	c	0	2	10	25
2	-1.4	0.28678	3.68042	52.01192	117.42231
	-1.3	0.16516	1.76508	24.81087	87.85212
	-1.2	0.06551	0.82382	13.31012	47.90199
	-1.1	0.01001	0.01121	6.71252	22.71814
	-1.0	0.00408	4.10×10^{-6}	0	0
	-0.9	0.00911	0.00919	4.78188	39.90123
	-0.8	0.05017	0.64912	11.68091	76.73219
	-0.6	0.16531	2.30843	42.09454	303.88752
	-0.4	0.32229	4.62449	84.93014	633.03826
	-0.2	0.50244	7.33061	135.25419	1020.91425
	0	0.69276	10.21722	189.09218	1436.21172
15	-1.4	0.00551	0.01822	0.14182	0.78691
	-1.3	0.00321	0.00981	0.09821	0.55653
	-1.2	0.00137	0.00423	0.03276	0.18301
	-1.1	0.00099	0.00124	0.00721	0.04540
	-1.0	0.00018	0	0	0
	-0.9	0.00089	0.00093	0.00632	0.06711
	-0.8	0.00111	0.00352	0.02789	0.28562
	-0.6	0.00035	0.01262	0.10271	0.57725
	-0.4	0.00069	0.02584	0.21152	1.19164
	-0.2	0.01081	0.04146	0.34261	1.93263
	0	0.01502	0.05852	0.48472	2.73639
25	-1.4	0.00199	0.00557	0.03442	0.16491
	-1.3	0.00092	0.00322	0.01284	0.09652
	-1.2	0.00050	0.00128	0.00797	0.03852
	-1.1	0.00012	0.00045	0.00059	0.00345
	-1.0	0.00007	0	0	0
	-0.9	0.00009	0.00077	0.00043	0.00256
	-0.8	0.00041	0.00165	0.00676	0.03307
	-0.6	0.00128	0.00384	0.02524	0.12229
	-0.4	0.00251	0.00796	0.05207	0.25287
	-0.2	0.00395	0.01286	0.08444	0.41087
	0	0.00549	0.01823	0.11947	0.58229

Table 8.2 (cont'd)

v	c	λ_2			
		0	2	10	25
50	-1.4	0.00050	0.00122	0.00582	0.02236
	-1.3	0.00035	0.00077	0.00313	0.01112
	-1.2	0.00012	0.00031	0.00139	0.00517
	-1.1	0.00008	0.00012	0.00092	0.00087
	-1.0	0.00002	0	0	0
	-0.9	0.00006	0.00011	0.00087	0.00063
	-0.8	0.00010	0.00024	0.00116	0.00438
	-0.6	0.00032	0.00086	0.00429	0.01682
	-0.4	0.00063	0.00177	0.00888	0.03477
	-0.2	0.00100	0.00286	0.01441	0.05658
	0	0.00139	0.00403	0.02043	0.08015

Table 8.3

The optimal critical value when estimating σ^2 using the ML component according to the criterion of minimum average relative risk

			$\lambda_2=0$		$\lambda_2=2$		$\lambda_2=10$		$\lambda_2=25$	
n	k	v	c^A_{ML}	$A(c^A_{ML})$	c^A_{ML}	$A(c^A_{ML})$	c^A_{ML}	$A(c^A_{ML})$	c^A_{ML}	$A(c^A_{ML})$
20	2	18	0	0.00154	0	0	0	0	0	0
	5	15	-1.8	0.04017	0	0	0	0	0	0
	10	10	-3.4	0.22263	-1.6	0.07794	0	0	0	0
	15	5	-3.9	0.74453	-2.3	0.17356	0	0	0	0
30	2	28	0	6.82×10^{-4}	0	0	0	0	0	0
	5	25	-1.8	0.01758	0	0	0	0	0	0
	10	20	-3.6	0.08729	-1.9	0.03185	0	0	0	0
	15	15	-4.5	0.23109	-3.1	0.13799	0	0	0	0
	25	5	-5.3	1.17460	-3.4	0.81833	0	0	0	0
40	2	38	0	3.83×10^{-4}	0	0	0	0	0	0
	5	35	-1.9	0.00976	0	0	0	0	0	0
	10	30	-3.6	0.04745	-2.2	0.00716	0	0	0	0
	15	25	-4.6	0.11457	-2.4	0.06914	0	0	0	0
	25	15	-6.1	0.41217	-3.3	0.31972	-0.5	0.07174	0	0
	35	5	-6.7	1.44054	-4.5	1.14941	-1.8	0.21971	0	0
50	2	48	0	2.46×10^{-4}	0	0	0	0	0	0
	5	45	-1.9	0.00621	0	0	0	0	0	0
	10	40	-3.8	0.00912	-2.1	0.01061	0	0	0	0
	15	35	-4.7	0.02952	-3.6	0.04084	0	0	0	0
	25	25	-6.3	0.21555	-5.1	0.16816	-0.8	0.07651	0	0
	35	15	-7.3	0.56817	-5.9	0.47893	-2.4	0.12607	0	0
	45	5	-7.9	0.75631	-6.2	0.55575	-3.7	0.29653	0	0
80	2	78	-0.1	9.66×10^{-5}	0	0	0	0	0	0
	5	75	-2.0	0.00241	0	0	0	0	0	0
	10	70	-3.8	0.00733	-2.1	0.00645	0	0	0	0
	15	65	-4.8	0.02413	-2.5	0.01281	0	0	0	0
	25	55	-6.5	0.06624	-3.9	0.06568	-1.3	0.00287	0	0
	35	45	-7.7	0.13702	-5.2	0.09718	-3.8	0.03491	0	0
	45	35	-8.8	0.25609	-6.6	0.17261	-5.0	0.10126	0	0
	55	25	-9.6	0.47290	-7.1	0.35168	-5.5	0.21690	-0.9	0.11234
	65	15	-10.2	0.94010	-8.3	0.61872	-5.7	0.43524	-1.3	0.29832
	75	5	-10.3	2.48450	-9.2	1.21242	-5.8	0.85549	-2.2	0.62183

Table 8.4

The optimal critical value when estimating σ^2 using the LS component according to the criterion of mini-max regret

a) $\lambda_2 = 0$

v	c_{LS}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
2	-1.09192	0.00442	-0.77748	2.10961	-0.39223
5	-1.11706	0.00134	-0.70138	1.72046	-0.32481
10	-1.12893	0.00043	-0.66868	1.59890	-0.29797
25	-1.13805	0.00008	-0.64635	1.48828	-0.28076
30	-1.13894	0.00006	-0.64381	1.48102	-0.27880
35	-1.13861	0.00004	-0.64197	1.45781	-0.27739
40	-1.14009	0.00003	-0.64059	1.44466	-0.27633
45	-1.14073	0.00003	-0.63941	1.44942	-0.27551
50	-1.14079	0.00002	-0.63864	1.48115	-0.27485
55	-1.14130	0.00002	-0.63783	1.46016	-0.27430
60	-1.14151	0.00002	-0.63723	1.45661	-0.27385
65	-1.14178	0.00001	-0.63669	1.41730	-0.27347
70	-1.14196	0.00001	-0.63625	1.45632	-0.27313
75	-1.14173	9.8×10^{-6}	-0.63605	1.35582	-0.27285
80	-1.14238	8.6×10^{-6}	-0.63552	1.47930	-0.27260

b) $\lambda_2 = 2$

v	c_{LS}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
2	-1.1323	6.5×10^{-6}	-2.71204	2.55231	-1.00020
5	-1.0872	1.3×10^{-8}	-2.66451	2.54312	-1.00009
10	-1.0643	4.2×10^{-9}	-2.53415	2.54107	-1.00004
25	-1.0241	0	-2.44911	2.53850	-1.00001
30	-1.0019	0	-2.34124	2.53109	-1.00000
35	-1.0004	0	-2.31790	2.52867	-1.00000
40	-1.0000	0	-2.30654	2.52111	-1.00000
45	-1.0000	0	-2.29172	2.51910	-1.00000
50	-1.0000	0	-2.28987	2.51871	-1.00000
55	-1.0000	0	-2.28143	2.51800	-1.00000
60	-1.0000	0	-2.27992	2.51717	-1.00000
65	-1.0000	0	-2.27921	2.51692	-1.00000
70	-1.0000	0	-2.27858	2.51604	-1.00000
75	-1.0000	0	-2.27812	2.51575	-1.00000
80	-1.0000	0	-2.27799	2.51502	-1.00000

where λ_1^L and λ_1^U are the values of λ_1 at which $Reg(\lambda, c^{MX})$ attains a maximum in the region $\lambda_1 \le \lambda_1^*$ and $\lambda_1 > \lambda_1^*$ respectively, and REG^* is the value of the regret function corresponding to λ_1^L and λ_1^U.

Table 8.5

The optimal critical value when estimating σ^2 using the MM component according to the criterion of mini-max regret

a) $\lambda_2 = 0$

v	c_{MM}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
2	-1.50612	0.06883	0.68826	2.74921	1.21341
5	-1.64195	0.02394	0.69001	2.46294	1.17325
10	-1.70241	0.00846	0.69092	2.32469	1.14962
25	-1.74401	0.00172	0.69116	2.22674	1.13051
30	-1.74891	0.00123	0.69117	2.21537	1.12808
35	-1.75239	0.00092	0.69119	2.21117	1.12629
40	-1.75511	0.00072	0.69120	2.20283	1.12495
45	-1.75725	0.00058	0.69121	2.20257	1.12385
50	-1.75874	0.00047	0.69122	2.19003	1.12305
55	-1.76030	0.00039	0.69126	2.18848	1.12225
60	-1.76147	0.00033	0.69127	2.18521	1.12166
65	-1.76247	0.00028	0.69129	2.18309	1.12116
70	-1.76331	0.00025	0.69130	2.18087	1.12069
75	-1.76485	0.00020	0.69134	2.17959	1.12042
80	-1.76571	0.00019	0.69138	2.17679	1.12023

b) $\lambda_2 = 2$

v	c_{MM}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
2	-0.80123	0.000877	-1.55621	2.10982	-0.92152
5	-0.87852	0.000128	-1.65412	1.99102	-0.95781
10	-0.95450	9.8×10^{-6}	-1.81201	1.90009	-0.99192
25	-0.96312	4.2×10^{-6}	-1.88512	1.82135	-1.00000
30	-0.96825	8.8×10^{-7}	-1.97152	1.77678	-1.00000
35	-0.97260	6.5×10^{-7}	-2.20991	1.65689	-1.00000
40	-0.97590	5.9×10^{-7}	-2.12876	1.55324	-1.00000
45	-0.97849	4.7×10^{-7}	-2.17182	1.41902	-1.00000
50	-0.98058	3.2×10^{-7}	-2.19005	1.36362	-1.00000
55	-0.98230	1.9×10^{-7}	-2.21017	1.22553	-1.00000
60	-0.98374	9.9×10^{-8}	-2.22001	1.17019	-1.00000
65	-0.98496	7.3×10^{-8}	-2.22991	1.10108	-1.00000
70	-0.98601	6.8×10^{-8}	-2.23154	0.98472	-1.00000
75	-0.98693	5.4×10^{-8}	-2.23912	0.90281	-1.00000
80	-0.98773	4.2×10^{-8}	-2.24728	0.88291	-1.00000

Table 8.6

The optimal critical value when estimating σ^2 using the ML component according to the criterion of mini-max regret

a) $\lambda_2 = 0$

n	k	v	c_{ML}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
20	2	18	-1.15947	0.00174	0.16353	1.49517	0.53730
	5	15	-2.58494	0.01811	1.73492	3.35189	2.30512
	10	10	-3.93490	0.07633	3.12302	5.14927	3.88897
	15	5	-4.78158	0.20032	3.99742	6.79720	5.00767
30	2	28	-1.17069	0.00078	0.16743	1.47507	0.53730
	5	25	-2.62188	0.00797	1.75384	3.29147	2.30513
	10	20	-4.02732	0.03117	3.19032	4.96748	3.88896
	15	15	-5.03071	0.06993	4.18372	6.24307	5.00767
	25	5	-6.32795	0.25455	5.45291	8.89936	6.71144
40	2	38	-1.17639	0.00044	0.16944	1.46303	0.53730
	5	35	-2.63890	0.00447	1.76250	3.26258	2.30513
	10	30	-4.06279	0.01695	3.21795	4.89826	3.88896
	15	25	-5.09459	0.03630	4.24608	6.10243	5.00766
	25	15	-6.64170	0.10436	5.73464	8.12175	6.67114
	35	5	-7.61737	0.28011	6.60589	10.00603	8.06414
50	2	48	-1.17966	0.00029	0.17062	1.45641	0.53730
	5	45	-2.64872	0.00285	1.76761	3.24666	2.30513
	10	40	-4.08175	0.01065	3.23422	4.86384	3.88897
	15	35	-5.12482	0.02231	4.27712	6.03405	5.00766
	25	25	-6.72003	0.05935	5.83143	7.91767	6.71144
	35	15	-7.93612	0.12575	6.96041	9.63365	8.06414
	45	5	-9.73430	0.22347	7.67833	12.50009	9.22081
80	2	78	-1.18462	0.00011	0.17237	1.45471	0.53730
	5	75	-2.66286	0.00111	1.77728	3.21882	2.30513
	10	70	-4.10764	0.00406	3.25609	4.81397	3.88897
	15	65	-5.16283	0.00826	4.31878	5.95334	5.00766
	25	55	-6.79340	0.02032	5.93361	7.72473	6.71144
	35	45	-8.09277	0.03792	7.19589	9.17566	8.06414
	45	35	-9.19527	0.06317	8.24035	10.48232	9.22081
	55	25	-10.14930	0.10101	9.10672	11.75273	10.24787
	65	15	-10.94241	0.10565	9.76910	13.15844	11.18105
	75	5	-11.24325	0.33208	9.95310	15.62321	12.04216

Table 8.6 (cont'd)

b) $\lambda_2 = 2$

n	k	v	c^{MX}_{ML}	REG^*	λ^L_1	λ^U_1	λ^*_1
20	2	18	0	0	/	/	/ [5]
	5	15	0	0	/	/	/
	10	10	-2.40995	0.03548	2.13244	3.66388	2.69107
	15	5	-3.00712	0.13116	3.36938	5.42503	4.13648
30	2	28	0	0	/	/	/
	5	25	0	0	/	/	/
	10	20	-2.66986	0.01409	2.14332	3.61301	2.69107
	15	15	-3.74003	0.04694	3.45421	5.17945	4.13648
	25	5	-4.45357	0.18905	5.07431	7.12918	6.08708
40	2	38	0	0	/	/	/
	5	35	0	0	/	/	/
	10	30	-2.77662	0.00749	2.14737	3.29742	2.69107
	15	25	-3.96077	0.02394	3.48106	5.10793	4.13648
	25	15	-5.36968	0.08618	5.22905	7.32101	6.08708
	35	5	-5.34612	0.23562	6.29889	9.55021	7.55212
50	2	48	0	0	/	/	/
	5	45	0	0	/	/	/
	10	40	-2.83502	0.00463	2.14845	3.59093	2.69107
	15	35	-4.06811	0.01449	3.49502	5.07651	4.13648
	25	25	-5.67684	0.04875	5.29330	7.17705	6.08708
	35	15	-6.61410	0.11257	6.55631	8.95416	7.55212
	45	5	-6.36879	0.24112	7.37377	10.98217	8.77643
80	2	78	0	0	/	/	/
	5	75	0	0	/	/	/
	10	70	-2.91535	0.00172	2.14888	3.57867	3.00004
	15	65	-4.20334	0.00522	3.50836	5.03727	4.13648
	25	55	-5.97155	0.01630	5.36229	7.08708	6.08708
	35	45	-7.27358	0.03317	6.73871	8.59964	7.55211
	45	35	-8.29649	0.05767	7.85480	9.96420	8.77642
	55	25	-9.06071	0.09430	8.77795	11.25781	9.84989
	65	15	-9.41389	0.15531	9.49804	12.64013	10.81744
	75	5	-8.23625	0.29171	9.82719	14.66722	11.70532

[5] When $c^{MX}_{ML} = 0$, $\tilde{\sigma}^2_{ML}$ strictly dominates both σ^{**2}_{ML} and $\hat{\sigma}^2_{ML}$. Hence it is irrelevant to consider the values of λ^L_1, λ^U_1 and λ^*_1.

Table 8.6 (cont'd)

b) λ_2 = 10

n	k	v	c_{ML}^{MX}	REG^*	λ_1^L	λ_1^U	λ_1^*
20	2	18	0	0	/	/	/
	5	15	0	0	/	/	/
	10	10	0	0	/	/	/
	15	5	0	0	/	/	/
30	2	28	0	0	/	/	/
	5	25	0	0	/	/	/
	10	20	0	0	/	/	/
	15	15	0	0	/	/	/
	25	5	-1.11639	0.16005	1.88691	3.01225	2.30513
40	2	38	0	0	/	/	/
	5	35	0	0	/	/	/
	10	30	0	0	/	/	/
	15	25	0	0	/	/	/
	25	15	-1.66726	0.00792	1.87605	3.02973	3.00000
	35	5	-2.22912	0.06891	4.34497	5.96383	5.00766
50	2	48	0	0	/	/	/
	5	45	0	0	/	/	/
	10	40	0	0	/	/	/
	15	35	0	0	/	/	/
	25	25	-1.91833	0.00463	1.86510	3.04545	2.30513
	35	15	-3.30540	0.03850	4.37641	5.88987	5.00766
	45	5	-2.95305	0.10490	5.87701	7.87666	6.71144
80	2	78	0	0	/	/	/
	5	75	0	0	/	/	/
	10	70	0	0	/	/	/
	15	65	0	0	/	/	/
	25	55	-2.23807	0.00154	1.84328	3.08277	2.30513
	35	45	-4.26563	0.01201	4.40072	5.83608	5.00766
	45	35	-4.32453	0.00390	3.00010	7.02753	3.00002
	55	25	-7.98913	0.00980	7.93857	8.06413	8.06413
	65	15	-5.96720	0.09661	8.24877	10.49699	9.22081
	75	5	-4.46383	0.15901	9.02293	11.90428	10.24787

APPENDIX 8B

Figure 8.1 : $\rho(\tilde{\sigma}^2_{MM}, \sigma^2)$ for n = 50, k = 25 and λ_2 = 0 with c = c^{MX} and c = c^{A}

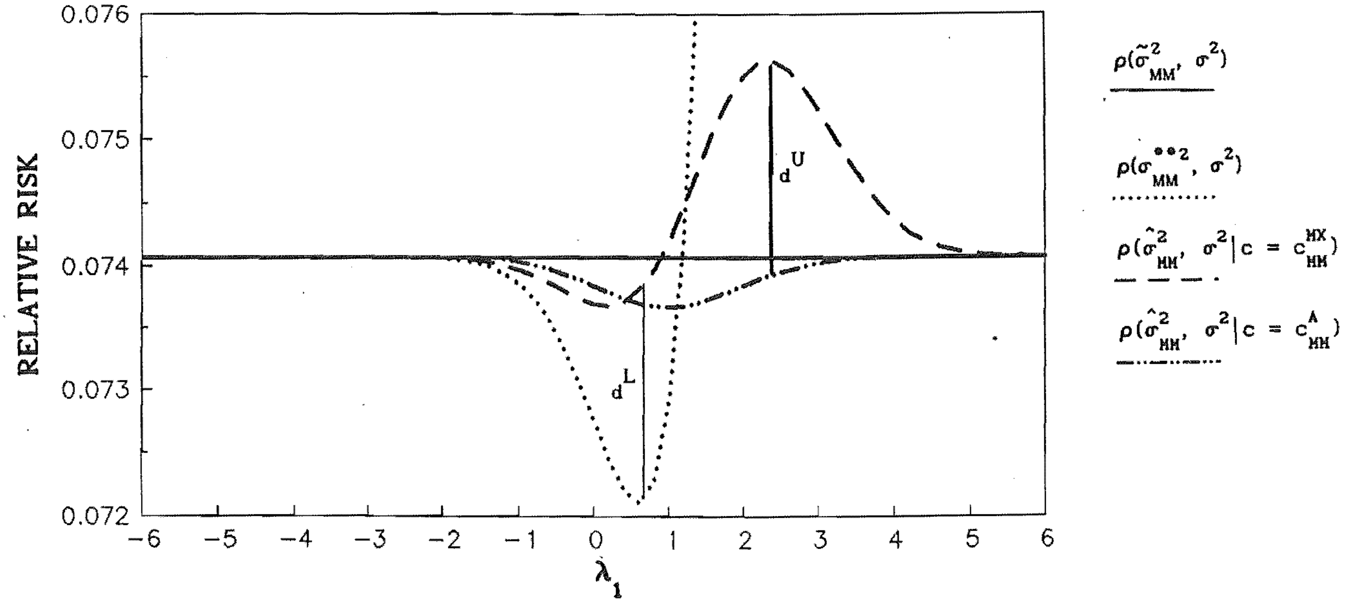


Figure 8.2 : $\rho(\tilde{\sigma}^2_{MM}, \sigma^2)$ for n = 80, k = 5 and λ_2 = 0 with c = c^{MX} and c = c^{A}

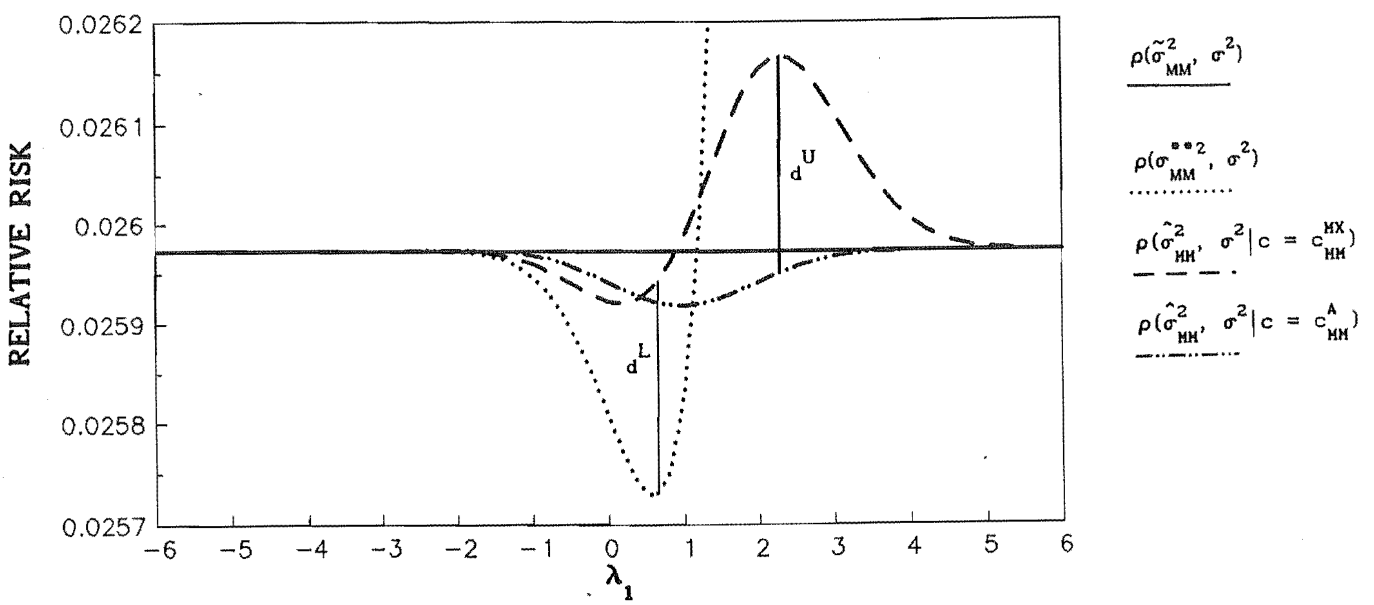


Figure 8.3 : $\rho(\tilde{\sigma}^2_{ML}, \sigma^2)$ for n = 20, k = 5 and $\lambda_2 = 0$ with $c = c^{MX}$ and $c = c^{A}$

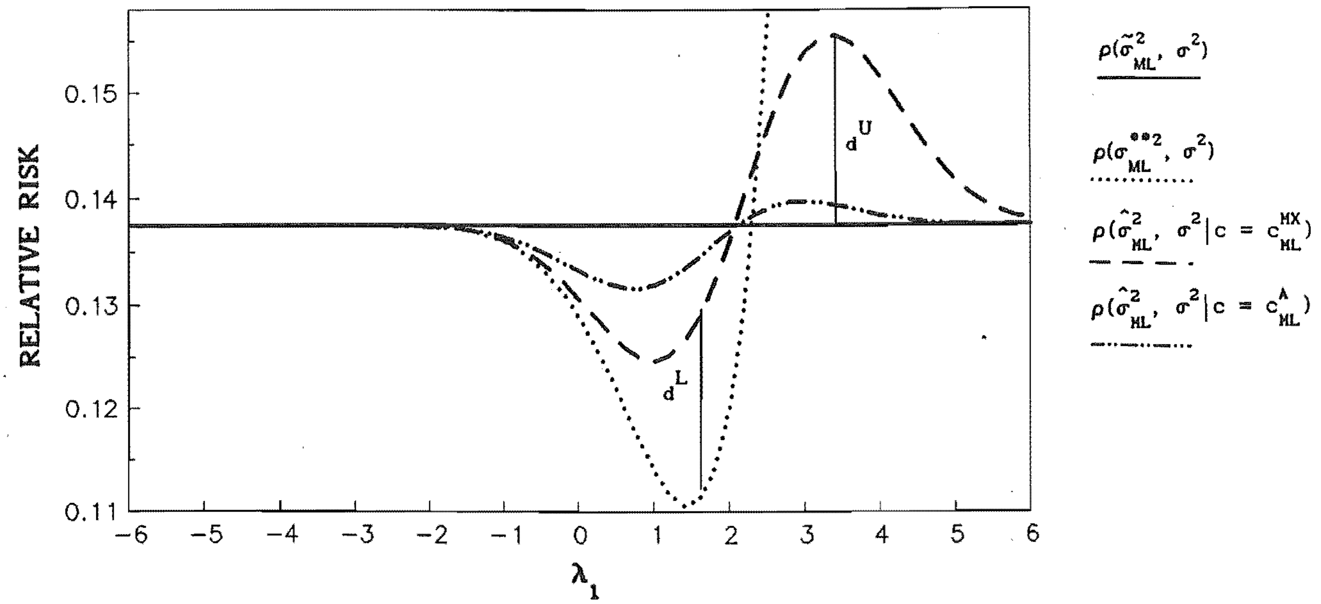


Figure 8.4 : $\rho(\tilde{\sigma}^2_{ML}, \sigma^2)$ for n = 50, k = 35 and $\lambda_2 = 0$ with $c = c^{MX}$ and $c = c^{A}$

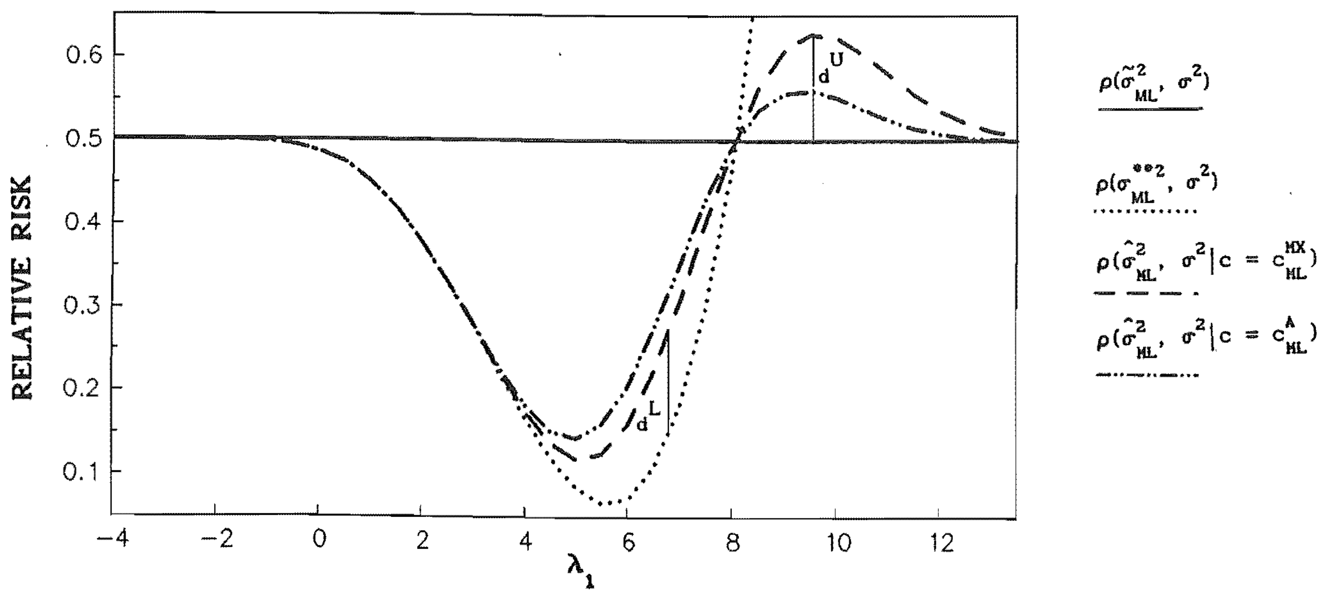


Figure 8.5 : $\rho(\tilde{\sigma}^2_{ML}, \sigma^2)$ for n = 20, k = 5 and λ_2 = 2 with $c = c^{MX}$ and $c = c^{A}$

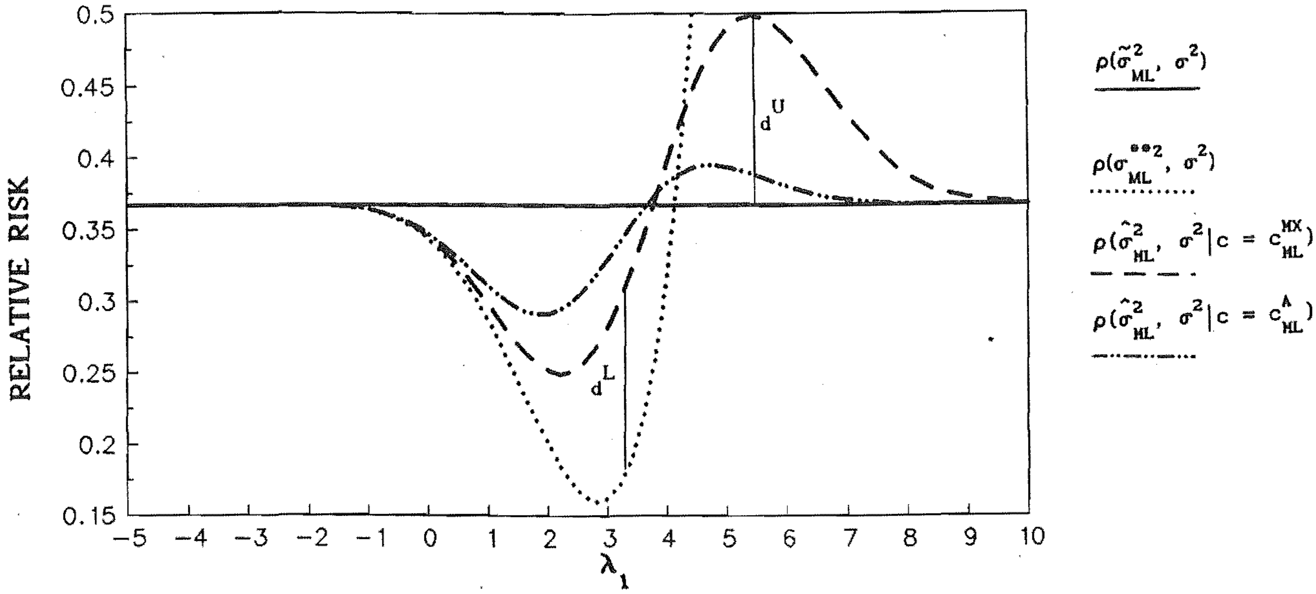
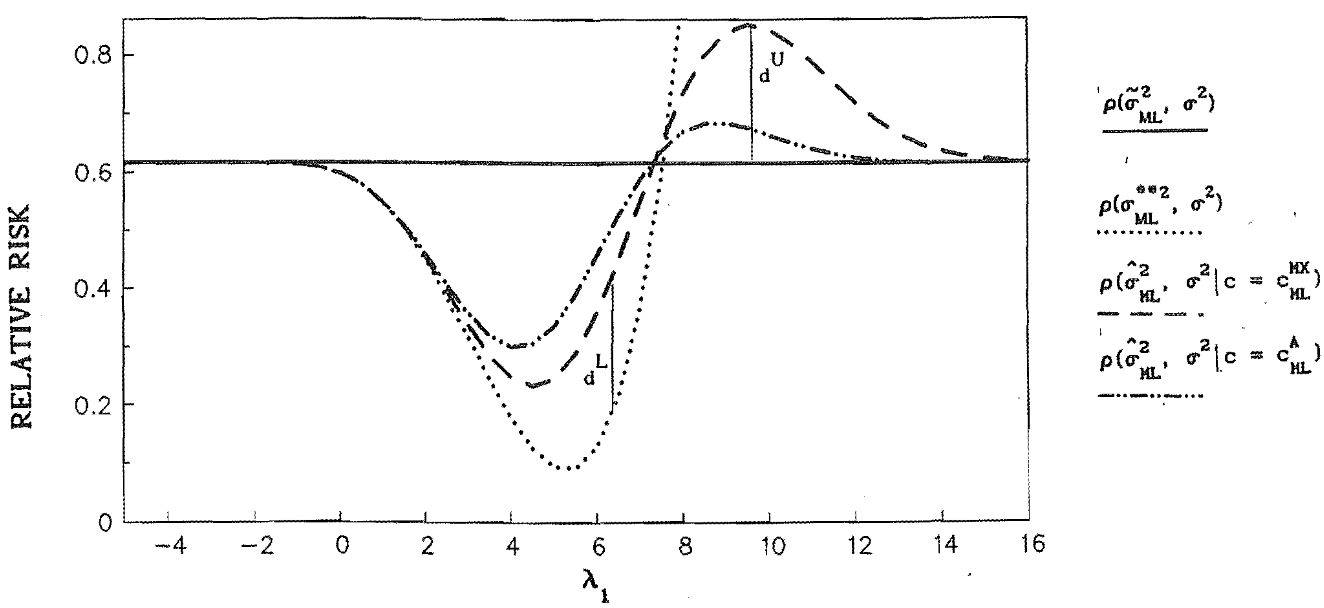


Figure 8.6 : $\rho(\tilde{\sigma}^2_{ML}, \sigma^2)$ for n = 50, k = 35 and λ_2 = 10 with $c = c^{MX}$ and $c = c^{A}$



CHAPTER NINE
SUMMARY AND CONCLUDING REMARKS

This thesis is directed towards expanding our knowledge on the finite sample properties of the inequality restricted and pre-test estimators in the linear regression model. To a large extent this investigation is motivated by the relative silence in the literature about many important issues regarding the statistical properties of these estimators. In particular, questions relating to the properties of these estimators for the error variance in small samples, the choice of an optimal critical value of the pre-test, and the effects of model mis-specification on the known properties of the inequality restricted and pre-test estimators have not been studied previously.

Throughout the analysis, we have chosen to focus on the single linear inequality constraint case in order to keep our results tractable. Arguably, this is also the type of inequality prior information that is most commonly encountered in practice : often in regression analysis, a researcher knows, or is at least willing to stipulate that a regression coefficient is larger or smaller than a known constant. The direction of the linear inequality is not crucial, but the assumption that there is only one constraint is essential for our results. We shall return to this point later in this chapter.

In Chapter 4 of the thesis, we examined the effects of mis-specifying the linear model, through the omission of relevant regressors, on the sampling performance of the inequality restricted and pre-test estimators for the regression prediction vector. The exact finite sample risks of these estimators were derived and evaluated, and their properties were contrasted with the situation when there is no mis-specification in the model. Our results show that underfitting the model has both quantitative and qualitative implications for the risk functions of the inequality restricted and pre-test estimators. Once the realistic possibility of omitted regressors is allowed for, there is no guarantee that the use of valid prior information will result in a risk improvement over the conventional maximum likelihood estimator. The risk of the inequality pre-test estimator also has the potential to be infinitely greater than the risk of the conventional maximum likelihood estimator when the degree of model mis-specification is serious. These results suggest that the findings of earlier works which assume that the constraint is in an exact equality form (for instance, Mittelhammer (1984), Giles (1990, 1991b)) qualitatively carry over to the case in which the restriction exists as a linear inequality.

The choice of an optimal critical value for the pre-test of an inequality restriction when estimating E(y) is the theme of Chapter 5 of this thesis. The optimality criteria adopted there are those of mini-max regret and minimum average relative risk. It is shown that when there is no excluded regressor, the mini-max regret critical value is roughly constant irrespective of the model's degrees of freedom. This provides a useful rule of thumb for applied researchers when testing for an inequality restriction in regression analysis. However, our results also show that this simple rule can be dangerously mis-leading when there is possible mis-specification in the regressor matrix, in which case the mini-max regret critical value varies according to the degrees of freedom, and the mini-max regret critical value can differ considerably from that obtained under the maintained assumption of a well specified model. Again, these results are qualitatively consistent with those reported in the literature for the case in which the restriction holds as a strict equality (Sawa and Hiromatsu (1973), Brook (1972, 1976) and Giles *et al.* (1992)). If the objective function is to minimize average relative risk, then our results suggest that one should always ignore the prior information, regardless of the degree of mis-specification in the design matrix. When there are no excluded regressors, this result again concurs qualitatively with that obtained when the *a priori* restriction holds as a strict equality (Toyoda and Wallace (1976)).

Chapter 6 of the thesis investigates the sampling properties of several estimators of the error variance which take into account the inequality restriction imposed on the regression coefficient vector in the standard linear model. To the best of the writer's knowledge, no other studies have analysed the properties of the estimators of the error variance within this context. We have derived and numerically evaluated the risk functions of a general family of estimators which comprise the ML, LS and MM component estimators as special members. For all three components that we have considered, our results show that when the inequality restriction is valid or nearly so, from the standpoint of minimizing estimator risk, it is always better to impose the restriction. When using the LS or MM component estimators, there is always a class of inequality pre-test estimators that strictly dominate the unrestricted estimator irrespective of the model's degrees of freedom. This feature is not noted when estimating E(y), but is consistent with the results reported in the literature when one is estimating σ^2 and the restriction holds as a strict equality (Ohtani (1988), Giles (1990, 1991a)).

Although this class of pre-test estimators is dominated by the inequality restricted estimator in the space where the restriction is true or close to being true, the degree of dominance is typically very slight. When using the ML component, depending on the choice of critical value, $\hat{\sigma}^2_{ML}$ dominates the unrestricted estimator over the entire range of the parameter space only when the model has a sufficiently large number of regression coefficients. Out of the three component estimators that we have considered, in terms of minimizing estimator risk, it is always preferable to use the estimator based on the minimum mean squared error component, other things being equal.

The results obtained in Chapter 6 form the motivation for the investigation discussed in Chapter 7, where the effects of omitting relevant regressors on the sampling performance of the inequality restricted and pre-test estimators of the error variance are analysed. Our results suggest that if the model is underfitted and the researcher is using the LS or MM components, then it is generally preferable to pre-test with $c = c^*$ than to ignore the inequality restriction or to impose the restriction indiscriminately, where c^* is equal to $-\sqrt{v/(v+2)}$ (for MM) or -1 (for LS). If, however, one is estimating based on the principle of maximum likelihood and the degree of model mis-specification is serious, then our results suggest that we should always ignore the prior information, even if it is perfectly correct.

We also tabulate optimal critical values of the pre-test when estimating σ^2 for the case where no strictly dominating estimator exists. This is considered in Chapter 8. Our results show that when using the LS or MM components, the optimal critical values according to the criterion of minimum average relative risk is always -1 for the LS case and $-\sqrt{v/(v+2)}$ for the MM case. This result is not surprising, given that the pre-test estimators corresponding to these critical values are the minimum risk estimators over almost the entire λ_1 range. When the criterion of mini-max regret is adopted, the optimal critical values are roughly constant across different values of v. Alternatively, when using the method of maximum likelihood, then the optimal critical value varies with both the number of observations and regression coefficients in model. Generally speaking, when there are no excluded regressors, both $|c^{MX}_{ML}|$ and $|c^{A}_{ML}|$ increase with the number of coefficients in model. Other things being equal, $|c^{MX}_{ML}|$ and $|c^{A}_{ML}|$ approach zero as λ_2 increases, reflecting the fact that $\tilde{\sigma}^2_{ML}$ strictly dominates all of the other estimators being considered when the model is sufficiently mis-specified.

Two major conclusions may be drawn from the thesis. First, common to both the problem of estimating the prediction vector and the error variance, when the model is underfitted, the use of perfectly correct valid information does not ensure a reduction in risk. This once again demonstrates that the credentials of certain traditional estimators often require strong and sometimes unsupportable assumptions regarding the underlying data generating process. Second, when estimating the scale parameter, we have shown that pre-testing can be the most advantageous strategy, and under certain conditions it is also robust to mis-specification of the regressor matrix. When estimating the prediction vector, however, specification error tends to work more favourably for the unrestricted estimator relative to the pre-test estimator. In practice, researchers rarely estimate the scale parameter and prediction vector separately. Given these results, it is unclear whether the risk gain from pre-testing in estimating the scale parameter can compensate the corresponding potential risk loss when estimating the prediction vector, especially when the degree of mis-specification is serious. This suggests that one should perhaps consider a joint risk function for estimating both the scale parameter and the prediction vector. This remains an interesting topic for further research.

The only type of model mis-specification that we have analysed in this thesis is that of excluding relevant regressors. The converse problem relating to the inclusion of irrelevant regressors was not studied rigorously. However, from the discussion in Appendix 7C, it appears that overfitting a model would not alter the usual risk comparisons among the estimators that have been considered. There are, of course, other types of mis-specification, such as those relating to the error distribution of the model, or the functional form of the constraint, that require attention. The impacts of these types of mis-specification on the properties of the inequality and pre-test estimators still remain to be investigated.

It should also be borne in mind that our results depend crucially on the assumption of a single linear inequality restriction. Given the findings of Thomson (1982), Judge and Yancey (1986), Yancey *et al.* (1989) and Judge *et al.* (1990), we do not expect our results to carry over to the multiple constraints situation in general. A major difficulty in extending our work to the multiple restrictions case lies in the fact that once there is more than one constraint, the correlation between the parameter estimates needs to be taken into account. Furthermore, the distributions of the test statistics for testing multiple inequality constraints are complicated functions of Chi-Squared random variables. Setting σ^2 to a known constant would reduce the complexity involved in analysing the properties of the pre-test estimator for the coefficient or prediction vector, but this would not be very realistic and it would not allow us to carry out any sensible analysis of the properties of the estimators for the scale parameter. The degree of complexity in this analysis could also be reduced by assuming that the X and Z matrices are orthogonal to each other as well as to themselves, but this would induce a considerable loss of generality. Surely, further research in this area is required before the properties of the inequality restricted and pre-test estimators are fully understood in a more general situation.

Furthermore, given our result that the risk comparisons between the unrestricted, inequality restricted and inequality pre-test estimators are distorted once the possibility of omitted regressors is allowed for, it would be interesting to extend these comparisons to include the family of "Stein-like" inequality estimators introduced by Judge *et al.* (1984). They show that under the assumption of a well specified model, the inequality restricted estimator is dominated by the inequality James and Stein estimator, which is in turn dominated by the positive part Stein inequality restricted estimator. There is, however, evidence which suggests that when there are excluded regressors, the usual James-Stein estimator no longer dominates the unrestricted estimator (Mittelhammer (1984)) as in the case where the model is well specified, but the positive part Stein estimator continues to be the best choice among the unrestricted, James-Stein and the positive part Stein estimators (Ohtani (1992)), though this estimator is also unlikely to be admissible. Whether these findings qualitatively carry over to the situation where there exists non-sample information of an inequality constraint form is still to be adequately examined.

In addition, given that we have now acquired knowledge on the sampling properties of the inequality restricted and pre-test estimators for σ^2, Stein like inequality estimators of σ^2 can be constructed, along the lines of Stein (1964), Judge *et al.* (1984) and Ohtani (1988). The specification of such estimators is interesting because there is a large body of literature which suggests that Stein type estimators and their variants often lead to a uniform improvement over the traditional estimators in terms of risk under quadratic loss in the context of various other problems. It would be interesting to know whether such impressive credentials for the Stein type estimator can be extended to wider situations that have so far received little or no attention in the literature. This remains for further research.

Also, throughout this thesis, the sampling performance of estimators are evaluated using the risk under squared error loss measure. The appeal of this measure lies in its ease of use and the fact that it incorporates the bias-variance trade-off. In the context of other pre-test problems, Giles (1992a), and Giles and Giles (1991, 1992, 1993b) reassess the sampling performance of certain pre-test estimators in terms of risk based on absolute error loss and asymmetric linex loss functions. The extension of these analyses to the problem of pre-testing of linear inequality restrictions remains a topic for future research.

We have also assumed throughout our investigation that the regression disturbances are normally distributed. Often economic data exhibit characteristics which suggest that they are not normally distributed.[1] In the context of pre-testing for exact equality restrictions and pre-testing for variance homogeneity, Giles (1990, 1991a, 1991b, 1992b and 1993) considers the properties of pre-test estimators under a wide family of non-normal disturbances. She finds that the wider error distribution assumption can have a substantial impact on the risk functions of the estimators of the scale parameter and the prediction vector (see also Wong and Giles (1991)). Again, the extension of these investigations to pre-testing inequality constraints remains to be done.

In conclusion, notwithstanding its limited scope, this thesis has expanded our knowledge on the sampling properties of the inequality restricted and pre-test estimators in linear regression. We have reached some general conclusions which should be of interest to econometricans in their applied work. Clearly, much remains to be done, and the results reported in this thesis have opened up avenues for further research in the general area of estimation with incomplete prior information.

---

[1] See Chapter 2, Section 2.6.2 for details.

REFERENCES

Albertson, K.V. (1991), "Pre-test estimation in a regression model with a mis-specified error covariance matrix", Discussion Paper No. 9115, Department of Economics, University of Canterbury.

Albertson, K.V. (1993), "Pre-test estimation in a regression model with a mis-specified error covariance matrix", Ph.D. thesis, University of Canterbury.

Bancroft, T.A. (1944), "On biases in estimation due to the use of preliminary tests of significance", *Annals of Mathematical Statistics* 15, 190-204.

Bartle, R.G. (1966), *The Elements of Integration*, John Wiley and Sons, New York.

Bock, M.E., T.A. Yancey and G.G. Judge (1973a), "The statistical consequences of preliminary test estimators in regression", *Journal of the American Statistical Association* 68, 109-116.

Bock, M.E., G.G. Judge and T.A. Yancey (1973b), "Some comments on estimation in regression after preliminary tests of significance", *Journal of Econometrics* 1, 191-200.

Box, G.E.P. (1952), "Multi-factor designs of first order", *Biometrika* 39, 49-57.

Brent, R.P. (1974), *Algorithms for Minimization Without Derivatives*, Prentice Hall, Englewood Cliffs.

Brook, R.J. (1972), "On the use of mini-max regret functions to set significance points in prior tests of estimation", Ph.D. thesis, North Carolina State University.

Brook, R.J. (1976), "On the use of a regret function to set significance points in prior tests of estimation", *Journal of the American Statistical Association* 71, 126-131.

Brook, R.J. and R.H. Fletcher (1981), "Optimal significance levels of prior tests in the presence of multicollinearity", *Communications in Statistics: Theory and Methods* A10, 1401-1413.

Brook, R.J. and V.K. Srivastava (1991), "Properties of interval constrained estimators", mimeo., Department of Statistics, Massey University.

Brook, R.J. and T.D. Wallace (1973), "A note on extraneous information in regression", *Journal of Econometrics* 1, 315-316.

Chiang, A.C. (1984), *Fundamental Methods of Mathematical Economics*, 3rd ed., New York: McGraw Hill.

Clarke, J.A. (1990), "Preliminary test estimation of the standard error of estimate in linear regression", *Economics Letters* 34, 27-32.

Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987a), "Estimating the error variance in regression after a preliminary test of restrictions on the coefficients", *Journal of Econometrics* 34, 293-304.

Clarke, J.A., D.E.A. Giles and T.D. Wallace (1987b), "Preliminary test estimation of the error variance in linear regression", *Econometric Theory* 3, 299-304.

Dantzig, G.B. and R.W. Cottle (1967), "Positive (Semi-) Definite Matrices and Mathematical Programming", in J. Abadie., ed., *Nonlinear Programming*, North-Holland: Amsterdam.

Davis, W.W. (1978), "Bayesian analysis of the linear regression model subject to linear inequality constraints", *Journal of the American Statistical Association* 73, 573-579.

De Doncker, E. (1978), "An adaptive exploration algorithm for automatic integration", *Signum Newsletter* 13, 2, 12-18.

Escobar, L.A. and B. Skarpness (1984), "A closed form for the least squares regression problem with linear inequality constraints", *Communications in Statistics: Theory and Methods* A13, 1127-1134.

Escobar, L.A. and B. Skarpness (1986), "The bias of the least squares estimator over interval constraints", *Economics Letters* 20, 331-335.

Escobar, L.A. and B. Skarpness (1987), "Mean square error and efficiency of the least squares estimator over interval constraints", *Communications in Statistics: Theory and Methods* A16, 397-406.

Fama, E.F. (1965), "The behaviour of stock market prices", *Journal of Business* 38, 34-105.

Farebrother, R.W. (1975), "Mini-max regret significance points for a preliminary test in regression analysis", *Econometrica* 43, 1005-1006.

Farebrother, R.W. (1978), "Estimating regression coefficients under conditional specification : Comment", *Communications in Statistics: Theory and Methods* A2, 193-196.

Farebrother, R.W. (1986), "Testing linear inequality constraints in the standard linear model", *Communications in Statistics: Theory and Methods* A15 , 7-31.

Farebrother, R.W. (1987), "A further remark on testing linear inequality constraints in the standard linear model", *Communications in Statistics: Theory and Methods* A16, 2003-2005.

Farebrother, R.W. (1988), "A survey of some recent econometric work on tests for inequality constraints", *Journal of Quantitative Economics* 4, 227-238.

Farebrother, R.W. (1990), "Simpler tests of linear inequality constraints in the standard linear model", paper presented to the 6th World Congress of the Econometric Society, Barcelona, Spain.

Gelfand, A.E. and D.K. Dey (1988), "Improved estimation of the disturbance variance in a linear regression model", *Journal of Econometrics* 39, 387-395.

Geweke, J. (1986), "Exact inference in the inequality constrained normal linear regression model", *Journal of Applied Econometrics* 1, 127-141.

Giles, D.E.A. (1986), "Preliminary test estimation in mis-specified regressions", *Economics Letters* 21, 325-328.

Giles, D.E.A. (1987), "Preliminary test estimation of mis-specified regression models", Discussion Paper No. 8703, Department of Economics, University of Canterbury.

Giles, D.E.A. (1992a), "Preliminary test estimation in regression under absolute error loss", Discussion Paper No. 9210, Department of Economics, University of Canterbury.

Giles, D.E.A. and J.A. Clarke (1989), "Preliminary test estimation of the scale parameter in a mis-specified regression model", *Economics Letters* 30, 201-205.

Giles, D.E.A. and J.A. Giles (1993b), "The exact risks of some pre-test and Stein-type regression estimators under balanced loss", mimeo., Department of Economics, University of Canterbury.

Giles, D.E.A., O. Lieberman and J.A. Giles (1992a), "The optimal size of a preliminary test of linear restrictions in a mis-specified regression model", *Journal of the American Statistical Association* 87, 1153-1157.

Giles, D.E.A., J.A. Giles and J.K. Wong (1992b), "Properties of a regression pre-test estimator in the presence of GARCH disturbances", mimeo., Department of Economics, University of Canterbury.

Giles, D.E.A. and V.K. Srivastava (1993), "The exact distribution of a least squares regression coefficient estimator after a preliminary t test", forthcoming in *Statistics and Probability Letters* 16, 59-64.

Giles, J.A. (1990), "Preliminary test estimation of a mis-specified linear model with spherically symmetric disturbances", Ph.D. thesis, University of Canterbury.

Giles, J.A. (1991a), "Pre-testing for linear restrictions in a regression with spherically symmetric disturbances", *Journal of Econometrics* 50, 377-398.

Giles, J.A. (1991b), "Pre-testing in a mis-specified regression model", *Communications in Statistics: Theory and Methods* A20, 3221-3238.

Giles, J.A. (1992b), "Estimation of the error variance after a preliminary test of homogeneity in a regression model with spherically symmetric disturbances", *Journal of Econometrics* 53, 345-361.

Giles, J.A. (1993), "Estimation of the scale parameter after a pre-test for homogeneity in a mis-specified regression model", forthcoming in *Communications in Statistics: Theory and Methods.*

Giles, J.A. and D.E.A. Giles (1991), "Preliminary test estimation of the regression scale parameter when the loss function is asymmetric", Discussion Paper No. 9104, Department of Economics, University of Canterbury.

Giles, J.A. and D.E.A. Giles (1992), "Estimation of the regression scale after a pre-test for homoscedasticity under LINEX loss", Discussion Paper No.9211, Department of Economics, University of Canterbury.

Giles, J.A. and D.E.A. Giles (1993a), "Pre-test estimation and testing in econometrics : Recent developments", forthcoming in *Journal of Economic Surveys.*

Giles, J.A. and O. Lieberman (1991), "The optimal size of a preliminary test for linear restrictions when estimating the regression scale parameter", *Economics Letters* 37, 25-30.

Gill, P.E., W. Murray and M.H. Wright (1981), *Practical Optimisation*, Academic Press: New York.

Goodnight, J. and T.D. Wallace (1972), "Operational techniques and tables for making weak MSE tests for restrictions in regression", *Econometrica* 40, 699-710.

Gourieroux, C., A. Holly and A. Monfort (1982), "Likelihood ratio, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters", *Econometrica* 50, 63-80.

Gun, A.M. (1965), "The use of a preliminary test for interactions in the estimation of factorial means", Series M436, Institute of Statistics, Raleigh, North Carolina.

Gun, A.M. (1967), "Use of preliminary test in the estimation of factorial means", *Calcutta Statistical Association Bulletin*, 16, 49-72.

Hasegawa, H. (1989a), "On some comparisons between Bayesian and sampling theoretic estimators of a normal mean subject to an inequality constraint", *Journal of the Japan Statistical Society* 19, 167-177.

Hasegawa, H. (1989b), "Bayesian estimator of the linear regression model with an interval constraint on coefficients", *Economics Letters* 7, 327-333.

Hasegawa, H. (1991), "The MSE of a pre-test estimator of the linear regression model with an interval constraint on coefficients", *Journal of the Japan Statistical Society* 21, 189-195.

Hillier, G. (1986), "Joint tests for zero restrictions on non-negative regression coefficients", *Biometrika* 73, 657-670.

Judge, G.G. and M.E. Bock (1978), *The Statistical Implications of Pre-test and Stein Rule Estimators in Econometrics*, North-Holland: Amsterdam.

Judge, G.G. and M.E. Bock (1983), "Biased Estimation", in Z. Grilliches and M. Intrilligator eds., *Handbook of Econometrics*, North Holland: Amsterdam.

Judge, G.G. and T. Takayama (1966), "Inequality restrictions in regression analysis", *Journal of the American Statistical Association* 61, 166-181.

Judge, G.G. and T.A. Yancey (1981), "Sampling properties of an inequality restricted estimator", *Economics Letters* 7, 327-333.

Judge, G.G. and T.A. Yancey (1986), *Improved Methods of Inference in Econometrics*, North-Holland: Amsterdam.

Judge, G.G., T.A. Yancey, M.E. Bock and R. Bohrer (1984), "The non-optimality of the inequality restricted estimator under squared error loss", *Journal of Econometrics* 25, 165-177.

Judge, G.G., T.A. Yancey and R. Bohrer (1988), "A comparison of the sampling performance of the positive Stein rule and two multivariate one-sided pre-test estimators under squared error loss", mimeo., Department of Agricultural and Resource Economics, University of California at Berkeley.

King, M.L. (1979), "Some aspects of statistical inference in the linear regression model", Ph.D. thesis, University of Canterbury.

King, M.L. and M.D. Smith (1986), "Joint one sided tests of linear regression coefficients", *Journal of Econometrics* 32, 367-383.

Klemn, R.J. and V.A. Sposito (1980), "Least squares solutions over interval restrictions", *Communications in Statistics: Simulation and Computation* B9, 423-425.

Kodde, D.A. and F.C. Palm (1986), "Wald criteria for jointly testing equality and inequality restrictions", *Econometrica* 54, 1243-1248.

Larson, H.J. and T.A. Bancroft (1963), "Biases in prediction by regression for certain incompletely specified models", *Biometrika* 50, 391-401.

Liew, C.K. (1976), "Inequality constrained least squares estimation", *Journal of the American Statistical Association* 71, 746-751.

Lovell, M.C. and E. Prescott (1970), "Multiple regression with inequality constraints: Pre-testing bias, hypothesis testing and efficiency", *Journal of the American Statistical Association* 65, 913-925.

Malinvaud, E. (1985), *Statistical Methods of Econometrics*, 4th ed., North-Holland: Amsterdam.

Mandelbröt, B.B. (1963), "The variation of certain speculative prices", *Journal of Business* 36, 394-419.

Metha, J.S. (1972), "On utilising information from a second sample in estimating the scale parameter for a family of symmetric distributions", *Journal of the American Statistical Association* 67, 448-452.

Mittelhammer, R.C. (1984), "Restricted least squares, pre-test, OLS and Stein rule estimators: Risk comparisons under model mis-specification", *Journal of Econometrics* 25, 151-164.

Numerical Algorithm Group Ltd. (1991), "NAG Fortran Library Manual, Mark 15, Vol. 1", NAG Inc., U.S.A.

O'Hagan, A. (1973), "Bayes estimation of a convex quadrature", *Biometrika* 60, 565-571.

Ohtani, K. (1983), "Preliminary test predictor in the linear regression model including a proxy variable", *Journal of the Japan Statistical Society* 17, 81-89.

Ohtani, K. (1987), "The MSE of the least squares estimator over an interval constraint", *Economics Letters* 25, 351-354.

Ohtani, K. (1988), "Optimal levels of significance of a pre-test in estimating the disturbance variance after the pre-test for a linear hypothesis on coefficients in a linear regression", *Economics Letters* 28, 151-156.

Ohtani, K. (1991a), "Estimation of the variance in a normal population after the one sided pre-test for the mean", *Communications in Statistics: Theory and Methods* A20, 219-234.

Ohtani, K. (1991b), "Some sampling properties of the inequality constrained least square estimator in a linear regression model with a proxy variable", mimeo, Faculty of Economics, Kobe University.

Ohtani, K. (1991c), "Some sampling properties of the least squares estimator over an interval constraint when error terms have a multivariate t distribution", *Journal of the Japan Statistical Society* 21, 197-204.

Ohtani, K. (1992), "On pre-testing in estimation of the coefficient of determination in a linear regression model with fixed regressors", Discussion Paper No. 9205, Faculty of Economics, Kobe University.

Ohtani, K. and T. Toyoda (1978), "Mini-max regret critical values for a preliminary test in pooling variances", *Journal of the Japan Statistical Society* 8, 15-20.

Özcam, A. and G.G. Judge (1991), "Some risk results for a two stage pre-test estimator in the case of possible heteroskedasticity", *Journal of Econometrics* 48, 355-371.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press: New York.

Rogers, A.J. (1986a), "Modified Lagrange multipler tests for models with one-sided alternatives", *Journal of Econometrics* 31, 341-361.

Rogers, A.J. (1986b), "Wald, Likelihood ratio and infinite induced test statistics for joint one-sided hypotheses", *Economics Letters* 21, 261-264.

Rogers, A.J. (1988), "The Bahadur efficiency of tests of some joint hypothesis", *Journal of the American Statistical Society* 83, 257-264.

Rothenberg, T. (1968), "Estimation with inequality restrictions", Heverlee, Belgium : for Operations Research and Econometrics, Discussion Paper No. 6825.

Savin, N.E. (1984), "Multiple Hypothesis Testing", in eds. Z. Grilliches and M.D. Intrilligator, *Handbook of Econometrics* Vol.2, North-Holland: Amsterdam.

Sawa, T. and T. Hiromatsu (1973), "Mini-max regret significance points for a preliminary test in regression analysis", *Econometrica* 45, 1293-1298.

Scheffé, H. (1953), "A method of judging all constraints in the analysis of variance", *Biometrika* 40, 87-104.

Shukla, N.D. (1979), "Estimation of a regression coefficient after two preliminary test of significance", *Metrika* 26, 182-193.

Srivastava, S.R. (1976), "A preliminary test estimator for the variance of a normal distribution", *Journal of the Indian Statistical Association* 14, 107-111.

Stein, C. (1964), "Inadmissibility of the usual estimator for the variance of a normal distribution with unknown means", *Annals of the Institute of Statistical Mathematics* 16, 155-160.

Theil, H. (1963), "On the use of incomplete prior information in regression analysis", *Journal of the American Statistical Association* 63, 558-572.

Theil, H. and A.S. Goldberger (1961), "On pure and mixed statistical estimation in economics", *International Economic Review* 2, 65-78.

Thomas, D.H. (1972), "Some contributions to radial probability distributions, statistics, and the operational calculi", Ph.D. thesis, Wayne State University.

Thomson, M. (1982), "Some results on the statistical properties of an inequality constrained least squares estimator in a linear model with two regressors", *Journal of Econometrics* 19, 215-231.

Thomson, M. and P. Schmidt (1982), "A note on the comparison of the mean square error of inequality constrained least squares and other related estimators", *Review of Economics and Statistics* 64, 174-176.

Toro-Vizcarrondo, C.E. (1968), "Multicollinearity and the mean square error criterion in multiple regression : A test and some sequential estimator comparisons", Ph.D. thesis, North Carolina State University.

Toro-Vizcarrondo, C.E. and T.D. Wallace (1968), "A test of the mean square error criterion for restrictions in linear regression", *Journal of the American Statistical Association* 63, 558-572.

Toyoda, T. and K. Ohtani (1978), "Optimal pre-testing procedure in regression - A minimum average risk approach", *Keizai Kenkyu* 29, 39-43.

Toyoda, T. and K. Ohtani (1986), "Testing equality between sets of coefficients after a pre-test for equality of disturbance variance in two linear regressions", *Journal of Econometrics* 31, 67-80.

Toyoda, T. and T.D. Wallace (1975), "Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test", *Journal of Econometrics* 3, 395-404.

Toyoda, T. and T.D. Wallace (1976), "Optimal critical values for pre-testing in regression", *Econometrica* 44, 365-375.

Varian, H.R. (1975), "A Bayesian approach to real estate assessment", in eds. S.E. Fienberg and A. Zellner, *Studies in Bayesian Econometrics and Statistics : in Honour of Leonard J. Savage, Vol. I*, Amsterdam: North Holland.

Wallace, T.D. (1972), "Weaker criteria and tests for linear restrictions in regression", *Econometrica* 40, 689-698.

Wallace, T.D. (1977), "Pre-test estimation in regression: A survey", *American Journal of Agricultural Economics* 59, 431-443.

Wallace, T.D. and C.E. Toro-Vizcarrondo (1968), "Tables for the mean square error test for exact linear restrictions in regression", *Journal of the American Statistical Association* 63, 1649-1665.

Wolak, F.A. (1987), "An exact test for multiple inequality and equality constraints in the linear model", *Journal of the American Statistical Association* 82, 782-793.

Wolak, F.A. (1988), "Duality in testing multivariate hypotheses", *Biometrika* 75, 611-615.

Wolak, F.A. (1989), "Testing inequality constraints in linear econometric models", *Journal of Econometrics* 41, 205-235.

Wolfe, P (1959), "The Simplex method for quadratic programming", *Econometrica* 27, 382-398.

Wong, J.K. and J.A. Giles (1991), "Optimal critical values of a preliminary test for linear restrictions in a regression model with multivariate Student t disturbances", Discussion Paper No. 9114, Department of Economics, University of Canterbury.

Yancey, T.A., R. Bohrer and G.G. Judge (1982), "Power function comparisons in inequality hypothesis testing", *Economics Letters* 9, 161-167.

Yancey, T.A., G.G. Judge and M.E. Bock (1981), "Testing multiple equality and inequality hypotheses in economics", *Economics Letters* 7, 249-255.

Yancey, T.A., G.G. Judge and R. Bohrer (1989), "Sampling performance of some joint one-sided preliminary test estimators under squared error loss", *Econometrica* 57, 1221-1228.

Zellner, A. (1961), "Linear regression with inequality constraints on the coefficients", Report No. 6109 of the International Centre for Management Science, Rotterdam.

Zellner, A. (1991), "Bayesian and Non-Bayesian estimation using balanced loss functions", Working Paper 91-109, Graduate School of Business, University of Chicago.